TRANSACTIONS
OF THE
AMERICAN INSTITUTE
OF
ELECTRICAL ENGINEERS

VOLUME 71
1952

PART II. APPLICATIONS AND INDUSTRY

PUBLISHED BY THE
AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS
33 WEST 39TH STREET
NEW YORK 18, N. Y., U. S. A.

Copyright 1952 by the American Institute of Electrical Engineers
Printed in U. S. A. by The Mack Printing Company, Easton, Pa.

Applications and Industry Index for 1952

1. Technical Subject Index

Actuators, Weight Analysis of Aircraft. Geyer, Treseder........118-26
Air Conditioning, Load Characteristics of Five All Electric Residences Using the Heat Pump for Year-Round. Sporn, Ambrose........353-60
(Air Transportation) A Dial-Reading Translator for Digital Machine Inputs. Klein, Gow........137-44; disc. 144
(Air Transportation) Arc Interruption Phenomena in a Magnetic Field at Altitude. Dallas........419-22
(Air Transportation) Consideration of Off-On-Modulated Reversing Clutch Servo Systems. Stuelpnagel, Dallas........406-10
(Air Transportation) Electrostatic Oscillation in Sensitive Contact Mechanisms. Stuelpnagel...397-401
Air Transportation, Progress in Applications of Electrical Engineering to. Fritz........261-9
(Air Transportation) Short-Circuiting Techniques to Deactivate a Generator After a Fault. Markowitz, Pobst........182-4
(Air Transportation) Stopping Devices for 400-Cycle Motors. Zahorsky........185-90; disc. 190
Airborne Electric and Electronic Equipment, A Review of the Factors Affecting the Temperature of. Lyons........423-5; disc. 425
Aircraft Actuators, Weight Analysis of. Geyer, Treseder........118-26
Aircraft Alternators, Automatic Synchronizing for. McCarty, Hills........454-7; disc. 458
Aircraft Cables, Continuous Current and Temperature Rise in. Schach........197-203
Aircraft, Continuous Current and Temperature Rise in Bundled Cables for. Schach, Kidwell.....376-84
Aircraft Cooling Systems for Rotating Electric Equipment, Study of. Martin........150-60
Aircraft D-C Systems, Stabilizing Elements in. Scorgie, Schaefer........387-96; disc. 396
Aircraft Distribution Systems, Impedance Data for 400-Cycle. Exner, Singer........410-19
Aircraft Electric Power Systems, Unit Equipments for. Finison........145-9
Aircraft Electric System, A New Circuit Breaker for. Beall, Reifschneider........109-12
Aircraft Electric Systems, Characteristics of Faults on. Larson........165-9
Aircraft Electric Systems, Distribution System Reliability of 28-Volt D-C. Carlson, Sherrard........113-17
Aircraft Electric Systems, Intermittent Faults in. Larson, Trbovich, Toomire........160-5
Aircraft Generators, Altitude Chamber Requirements for Evaluating Blast-Cooled. Scott........427-30; disc. 430
Aircraft Generators, Effects of Terminal Voltage, Load Current and Minimum Rotor Speed on the Weight of D-C. Scott........191-7
Aircraft Generators, Feeder-Fault Protection for D-C. Davis........430-5
Aircraft Radio Interference Measurements. Newman, Schwantes, Stahmann........211-15
Aircraft, Reduced Voltage D-C Motor Starting Controllers for. Finison, Andersen........126-33
Airplanes by Means of Electric Equipment, Improving the Dynamic Response of. Monroe........441-7; disc. 447
A-C Overcurrent Protective Device Settings for Industrial Plants, Selecting. Brightman........203-11
Alternators, Automatic Synchronizing for Aircraft. McCarty, Hills........454-7; disc. 458
Altitude, Arc Interruption Phenomena in a Magnetic Field at. Dallas........419-22
Altitude Chamber Requirements for Evaluating Blast-Cooled Aircraft Generators. Scott........427-30; disc. 430
Amplifier, Vacuum-Tube Techniques Applied to a Hydraulic. Baring........302-06
Amplitude Oscillation, Stabilization of a Servomechanism Subject to Large. Sherrard........312-24
Analogue Computers for Feedback Control Systems. Bruns........250-4
Analogue Methods for Optimum Servomechanism Design. Fickeisen, Stout........244-50; disc. 250
Analysis, A Block-Diagram Approach to Network. Stout........255-60
Analysis of Aircraft Actuators, Weight. Geye Treseder........118-26
Analysis of Feedback-Control Systems Containing Nonlinear Elements, Sinusoidal. Johnson.....169-181
Analysis of Sampled-Data Systems, The. Ragazzini, Zadeh........225-32; disc. 232
Anodes for Cathodic Protection of the Bottoms of Inactivated Ships, The Use of Graphite. Oliver........324-9
Anodes, Measurement of Cathodic Protection Currents from Sacrificial. Hayward, Wainwright....350-2
Apparatus for the Study of the Human Operator in a 1-Dimensional Closed-Loop Continuous Pursuit Task, An Electronic. Warren, Fitts, Clark........19-23
Application of Motors to Household Refrigeration Compressors. Packer........70-6; disc. 77
Applications of Electrical Engineering to Air Transportation, Progress in. Fritz........261-9
Applications, Tests of Electrostatic Controls for Hazardous Industrial. Beach........1-12; disc. 12
Approximate Transfer Function for the Analysis and Design of Pulsed Servos, An. Brown, Murphy........435-40; disc. 440
Arc Interruption Phenomena in a Magnetic Field at Altitude. Dallas........419-22
Automatic Synchronizing for Aircraft Alternators. McCarty, Hills........454-7; disc. 458

B

Blast-Cooled Aircraft Generators, Altitude Chamber Requirements for Evaluating. Scott........427-30; disc. 430
Block-Diagram Approach to Network Analysis, A. Stout........255-69
Breaker for Aircraft Electric System, A New Circuit. Beall, Reifschneider........109-12
Brightening with High-Frequency Rotating Equipment Induction Preheating of Electrolytic Tin Plate for Flow. Thomas........13-18
Bundled Cables for Aircraft, Continuous Current and Temperature Rise in. Schach, Kidwell....376-84
Bus Bar Design for High-Frequency Induction Heating. Williamson........66-9

C

Cables, Continuous Current and Temperature Rise in Aircraft. Schach........197-203
Cables for Aircraft, Continuous Current and Temperature Rise in Bundled. Schach, Kidwell....376-84
Cathodic Protection Currents from Sacrificial Anodes, Measurement of. Hayward, Wainwright...350-2
Cathodic Protection of the Bottoms of Inactivated Ships. The Use of Graphite Anodes for. Oliver....324-9
Century of Engineering, Electric Lighting in the First. Oetting........269-91
Chamber Requirements for Evaluating Blast-Cooled Aircraft Generators, Altitude. Scott........427-30; disc. 430
Characteristics of Faults on Aircraft Electric Systems. Larson........165-9
Characteristics of Five All Electric Residences Using the Heat Pump for Year-Round Air Conditioning, Load. Sporn, Ambrose........353-60
Chemical Plant, Electric Power Supply for a Large. Friel, Smith........83-9
Circuit Breaker for Aircraft Electric System, A New. Beall, Reifschneider........109-12
Circuiting Techniques to Deactivate a Generator After a Fault, Short-. Markowitz, Pobst........182-4
Closed-Loop Continuous Pursuit Task, An Electronic Apparatus for the Study of the Human Operator in a 1-Dimensional. Warren, Fitts, Clark........19-23
Closed-Loop Control System Transfer Functions by Inspection, The Writing of. Flannigan, Kirschbaum........90-4
Closed Loop Systems Using Curvilinear Squares to Predict Root Location, Synthesis of. Russell, Weaver........95-104
Clutch Servo Systems, Consideration of Off-On-Modulated Reversing. Stuelpnagel, Dallas.406-10
Communities, Co-ordinated Transportation for Large Metropolitan. Kearns........54-7; disc. 57
Component Control, Signal. Gimpel, Calvert..339-43
Compressors, Application of Motors to Household Refrigeration. Packer........70-6; disc. 77
Computers for Feedback Control Systems, Analogue. Bruns........250-4
Consideration of Off-On-Modulated Reversing Clutch Servo Systems. Stuelpnagel, Dallas........406-10
Contact Mechanisms, Electrostatic Oscillation in Sensitive. Stuelpnagel........397-401
Contracts, Railway Power. Martin........24-6
Control for Trolley Coaches, A New. Willby........63-5; disc. 65
Control, Signal Component. Gimpel, Calvert..339-43
Control System by Phase-Angle Loci, Synthesis of Feedback. Chu........330-9
Control System Transfer Functions by Inspection, The Writing of Closed-Loop. Flannigan, Kirschbaum........90-4
Control Systems, Analogue Computers for Feedback. Bruns........250-4
Control Systems with Dead-Time Lag or Distributed Lag by Root-Locus Method, Feedback. Chu........291-6
Controllers for Aircraft, Reduced Voltage D-C Motor Starting. Finison, Andersen........126-33
Controls for Hazardous Industrial Applications, Tests of Electrostatic. Beach........1-12; disc. 12
Cooling Systems for Rotating Electric Equipment, Study of Aircraft. Martin........150-60
Co-ordinated Transportation for Large Metropolitan Communities. Kearns........54-7; disc. 57
Core Memory, Lead Networks Utilizing a Saturable. Scorgie........297-302
Current and Temperature Rise in Aircraft Cables, Continuous. Schach........197-203
Current and Temperature Rise in Bundled Cables for Aircraft, Continuous. Schach, Kidwell....376-84
Currents from Sacrificial Anodes, Measurement of Cathodic Protection. Hayward, Wainwright........350-2
Curvilinear Squares to Predict Root Location, Synthesis of Closed Loop Systems Using. Russell, Weaver........95-104

D

Data for 400-Cycle Aircraft Distribution Systems, Impedance. Exner, Singer........410-19
Data Systems, The Analysis of Sampled-. Ragazzini, Zadeh........225-32; disc. 232
Deactivate a Generator After a Fault, Short-Circuiting Techniques to. Markowitz, Pobst........182-4
Dead-Time Lag or Distributed Lag by Root-Locus Method, Feedback Control Systems with. Chu........291-6
Design for High-Frequency Induction Heating, Bus Bar. Williamson........66-9
Devices for 400-Cycle Motors, Stopping. Zahorsky........185-90; disc. 190
Diagram Approach to Network Analysis, A Block-. Stout........255-60
Dial-Reading Translator for Digital Machine Inputs, A. Klein, Gow........137-44; disc. 144
D-C Aircraft Electric Systems, Distribution System Reliability of 28-Volt. Carlson, Sherrard.....113-17
D-C Aircraft Generators, Effects of Terminal Voltage, Load Current and Minimum Rotor Speed on the Weight of. Scott........191-7
D-C Aircraft Generators, Feeder-Fault Protection for. Davis........430-5
D-C Motor Starting Controllers for Aircraft, Reduced Voltage. Finison, Andersen........126-33
D-C Motor When Operated from Grid-Controlled Rectifiers, Increased Losses in a. Reiter, Ammerman........77-82
D-C Systems, Stabilizing Elements in Aircraft. Scorgie, Schaefer........387-96; disc. 396
Diesel-Electric Locomotives and Some Future Aspects of Electric Traction, Railway Electrification. Huldschiner........48-50
Diesel-Electric Locomotives, Maintenance and Operating Experiences—. Griffith, Pfeil.134-6; disc. 136
Diesel-Electric Switchers, 15 Years' Progress in the Design of Industrial. Barrell........215-19
Digital Machine Inputs, A Dial-Reading Translator for. Klein, Gow........137-44; disc. 144
Distribution System Reliability of 28-Volt D-C Aircraft Electric Systems. Carlson, Sherrard........113-17
Distribution Systems, Impedance Data for 400-Cycle Aircraft. Exner, Singer........410-19
(Dow Chemical Co.) Electric Power Supply for a Large Chemical Plant. Friel, Smith........83-9
Dynamic Response of Airplanes by Means of Electric Equipment, Improving the. Monroe........441-7; disc. 447

Vo
Pt
19

E

Earth as a Heat Source, Residential Heat Pump Experiments in Philadelphia—. Kidder, Neher...343–50
Effects of Terminal Voltage, Load Current and Minimum Rotor Speed on the Weight of D-C Aircraft Generators. Scott........191–7
Electric and Electronic Equipment, A Review of the Factors Affecting the Temperature of Airborne. Lyons........423–5; disc. 425
Electric Equipment, Study of Aircraft Cooling Systems for Rotating. Martin........150–60
Electric Glass Welding. Shaw........384–7
Electric Lighting in the First Century of Engineering. Oetting........269–91
Electric Locomotive for the Pennsylvania Railroad, A New. Gowans, Widell, Bredenberg........27–35; disc. 35
Electric Locomotives, Maintenance and Operating Experiences—Diesel-. Griffith, Pfeil........134–6; disc. 136
Electric Power Supply for a Large Chemical Plant. Friel, Smith........83–9
Electric Power Systems, Unit Equipments for Aircraft. Finison........145–9
Electric Spark Machining, Theory of. Williams........105–08; disc. 108
Electric System, A New Circuit Breaker for Aircraft. Beall, Reifschneider........109–12
Electric Systems, Characteristics of Faults on Aircraft. Larson........165–9
Electric Systems, Distribution System Reliability of 28-Volt D-C Aircraft. Carlson, Sherrard.....113–17
Electric Systems, Intermittent Faults in Aircraft. Larson, Trbovich, Toomire........160–5
Electrical Engineering to Air Transportation, Progress in Applications of. Fritz........261–9
Electrification, Diesel-Electric Locomotives, and Some Future Aspects of Electric Traction, Railway. Huldschiner........48–50
Electrification of Holland's Railway System. Candee........235–43; disc. 243
Electrolytic Tin Plate for Flow Brightening with High-Frequency Rotating Equipment, Induction Preheating of. Thomas........13–18
Electronic Apparatus for the Study of the Human Operator in a 1-Dimensional Closed-Loop Continuous Pursuit Task, An. Warren, Fitts, Clark........19–23
Electronic Equipment, A Review of the Factors Affecting the Temperature of Airborne Electric and. Lyons........423–5; disc. 425
Electrostatic Controls for Hazardous Industrial Applications, Tests of. Beach........1–12; disc. 12
Electrostatic Oscillation in Sensitive Contact Mechanisms. Stuelpnagel........397–401
Elements in Aircraft D-C Systems, Stabilizing. Scorgie, Schaefer........387–96; disc. 396
Elements, Sinusoidal Analysis of Feedback-Control Systems Containing Nonlinear. Johnson..169–181
Equipment, A Review of the Factors Affecting the Temperature of Airborne Electric and Electronic. Lyons........423–5; disc. 425
Equipment, Improving the Dynamic Response of Airplanes by Means of Electric. Monroe........441–7; disc. 447
Equipment, Induction Preheating of Electrolytic Tin Plate for Flow Brightening with High-Frequency Rotating. Thomas........13–18
Equipment, Study of Aircraft Cooling Systems for Rotating Electric. Martin........150–60
Equipments for Aircraft Electric Power Systems, Unit. Finison........145–9
Experiences—Diesel-Electric Locomotives, Maintenance and Operating. Griffith, Pfeil...134–6; disc. 136
Experiments in Philadelphia—Earth as a Heat Source, Residential Heat Pump. Kidder, Neher...343–50
Experiments in Philadelphia—Installation and Operating Experience, Residential Heat Pump. Harlow, Klapper........366–75
Experiments in Philadelphia—Suggested Possibilities for Practical Applications, Residential Heat Pump. Bary........361–5

F

Fault Protection for D-C Aircraft Generators, Feeder-. Davis........430–5
Fault. Short-Circuiting Techniques to Deactivate a Generator After a. Markowitz, Pobst........182–4
Faults in Aircraft Electric Systems, Intermittent. Larson, Trbovich, Toomire........160–5
Faults on Aircraft Electric Systems, Characteristics of. Larson........165–9
(Feedback Control) An Approximate Transfer Function for the Analysis and Design of Pulsed Servos. Brown, Murphy........435–40; disc. 440
(Feedback Control) Lead Networks Utilizing a Saturable Core Memory. Scorgie........297–302
(Feedback Control) Piecewise Linear Servomechanisms. Schwartz........401–05
(Feedback Control) Signal Component Control. Gimpel, Calvert........339–43
(Feedback Control) Stability Limits for Third-Order Servomechanisms. Higgins, Levinthal........459–66; disc. 466
Feedback Control System by Phase-Angle Loci, Synthesis of. Chu........330–9
Feedback Control Systems, Analogue Computers for. Bruns........250–4
Feedback Control Systems Containing Nonlinear Elements, Sinusoidal Analysis of. Johnson...169–181
Feedback Control Systems with Dead-Time Lag or Distributed Lag by Root-Locus Method. Chu........291–6
(Feedback Control) The Analysis of Sampled-Data Systems. Ragazzini, Zadeh.....225–32; disc. 232
(Feedback Control) Vacuum-Tube Techniques Applied to a Hydraulic Amplifier. Baring........302–06
Feedback to Improve the Transient Response of a Servomechanism, the Use of Nonlinear. Lewis........449–53; disc. 453
Feeder-Fault Protection for D-C Aircraft Generators. Davis........430–5
Flow Brightening with High-Frequency Rotating Equipment, Induction Preheating of Electrolytic Tin Plate for. Thomas........13–18
400-Cycle Aircraft Distribution Systems, Impedance Data for. Exner, Singer........410–19
Function for the Analysis and Design of Pulsed Servos, An Approximate Transfer. Brown, Murphy........435–40; disc. 440
Functions by Inspection, The Writing of Closed-Loop Control System Transfer. Flannigan, Kirschbaum........90–4

G

Generator After a Fault, Short-Circuiting Techniques to Deactivate a. Markowitz, Pobst........182–4
Generators, Altitude Chamber Requirements for Evaluating Blast-Cooled Aircraft. Scott........427–30; disc. 430
Generators, Effects of Terminal Voltage, Load Current and Minimum Rotor Speed on the Weight of D-C Aircraft. Scott........191–7
Generators, Feeder-Fault Protection for D-C Aircraft. Davis........430–5
Glass Welding, Electric. Shaw........384–7
Graphite Anodes for Cathodic Protection of the Bottoms of Inactivated Ships, The Use of. Oliver....324–9
Grid-Controlled Rectifiers, Increased Losses in a D-C Motor When Operated from. Reiter, Ammerman........77–82

H

Hazardous Industrial Applications, Tests of Electrostatic Controls for. Beach........1–12; disc. 12
Heat Pump Experiments in Philadelphia—Earth as a Heat Source, Residential. Kidder, Neher..343–50
Heat Pump Experiments in Philadelphia—Installation and Operating Experience, Residential. Harlow, Klapper........366–75
Heat Pump Experiments in Philadelphia—Suggested Possibilities for Practical Applications, Residential. Bary........361–5
Heat Pump for Year-Round Air Conditioning, Load Characteristics of Five All-Electrics Residences Using the. Sporn, Ambrose........353–60
Heating, Bus Bar Design for High-Frequency Induction. Williamson........66–9
High-Frequency Induction Heating, Bus Bar Design for. Williamson........66–9
High-Frequency Rotating Equipment, Induction Preheating of Electrolytic Tin Plate for Flow Brightening with. Thomas........13–18
Holland's Railway System, Electrification of. Candee........235–43; disc. 243
Horsepower of Synchronous and Induction Motors, Voltage Rating Versus. Miller........306–11
Household Refrigeration Compressors, Application of Motors to. Packer........70–6; disc. 77
Human Operator in a 1-Dimensional Closed-Loop Continuous Pursuit Task, An Electronic Apparatus for the Study of the. Warren, Fitts, Clark......19–23
Hydraulic Amplifier, Vacuum-Tube Techniques Applied to a. Baring........302–06

I

Ignitron-Rectifier Locomotive, The Pennsylvania Railroad. Whittaker, Hutchison......37–42; disc. 42
Impedance Data for 400-Cycle Aircraft Distribution Systems. Exner, Singer........410–19
Improving the Dynamic Response of Airplanes by Means of Electric Equipment. Monroe........441–7; disc. 447
Increased Losses in a D-C Motor When Operated from Grid-Controlled Rectifiers. Reiter, Ammerman........77–82
Induction Heating, Bus Bar Design for High-Frequency. Williamson........66–9
Induction Motors, Voltage Rating Versus Horsepower of Synchronous and. Miller........306–11
Induction Preheating of Electrolytic Tin Plate for Flow Brightening with High-Frequency Rotating Equipment. Thomas........13–18
Industrial Applications, Tests of Electrostatic Controls for Hazardous. Beach........1–12; disc. 12
(Industrial Control) Increased Losses in a D-C Motor When Operated From Grid-Controlled Rectifiers. Reiter, Ammerman........77–82
Industrial Diesel-Electric Switchers, 15 Years' Progress in the Design of. Barrell........215–19
Industrial Plants, Selecting A-C Overcurrent Protective Device Settings for. Brightman........203–11
Inputs, A Dial-Reading Translator for Digital Machine. Klein, Gow........137–44; disc. 144
Inspection, The Writing of Closed-Loop Control System Transfer Functions by. Flannigan, Kirschbaum........90–4
Interference Measurements, Aircraft Radio. Newman, Schwantes, Stahmann........211–15
Intermittent Faults in Aircraft Electric Systems. Larson, Trbovich, Toomire........160–5
Interruption Phenomena in a Magnetic Field at Altitude, Arc. Dallas........419–22

L

Lag or Distributed Lag by Root-Locus Method, Feedback Control Systems with Dead-Time. Chu........291–6
(Land Transportation) A New Control for Trolley Coaches. Willby........63–5; disc. 65
(Land Transportation) A New Electric Locomotive for the Pennsylvania Railroad. Gowans, Widell, Bredenberg........27–35; disc. 35
(Land Transportation) A New Wheel-Slip Protective Scheme. Smith........51–3
(Land Transportation) Co-ordinated Transportation for Large Metropolitan Communities. Kearns........54–7; disc. 57
(Land Transportation) Electrification of Holland's Railway System. Candee........235–43; disc. 243
(Land Transportation) 15 Years' Progress in the Design of Industrial Diesel-Electric Switchers. Barrell........215–19
(Land Transportation) Maintenance and Operating Experiences—Diesel-Electric Locomotives. Griffith, Pfeil........134–6; disc. 136
(Land Transportation) New Series Type Motor for Trolley Coaches. Woods, Steen........59–62
(Land Transportation) Railway Electrification, Diesel-Electric Locomotives, and Some Future Aspects of Electric Traction. Huldschiner........48–50
(Land Transportation) Railway Power Contracts. Martin........24–6
(Land Transportation) The Pennsylvania Railroad Ignitron-Rectifier Locomotive. Whittaker, Hutchison........37–42; disc. 42
Lead Networks Utilizing a Saturable Core Memory. Scorgie........297–302
Lighting in the First Century of Engineering, Electric. Oetting........269–91
Limits for Third-Order Servomechanisms, Stability. Higgins, Levinthal........459–66; disc. 466
Linear Servomechanisms, Piecewise. Schwartz........401–05
Load Characteristics of Five All-Electric Residences Using the Heat Pump for Year-Round Air Conditioning. Sporn, Ambrose........353–60
Load Current and Minimum Rotor Speed on the Weight of D-C Aircraft Generators, Effects of Terminal Voltage. Scott........191–7
Loci, Synthesis of Feedback Control System by Phase-Angle. Chu........330–9
Locomotive for the Pennsylvania Railroad, A New Electric. Gowans, Widell, Bredenberg........27–35; disc. 35
Locomotive, The Pennsylvania Railroad Ignitron-Rectifier. Whittaker, Hutchison...37–42; disc. 42
Locomotives and Some Future Aspects of Electric Traction, Railway Electrification, Diesel-Electric. Huldschiner........48–50
Locomotives, Maintenance and Operating Experiences—Diesel-Electric. Griffith, Pfeil....134–6; disc. 136
Locus Method, Feedback Control Systems with Dead-Time Lag or Distributed Lag by Root-. Chu........291–6
Loop Continuous Pursuit Task, An Electronic Apparatus for the Study of the Human Operator in a 1-Dimensional Closed-. Warren, Fitts, Clark..19–23
Loop Control System Transfer Functions by Inspection, The Writing of Closed-. Flannigan, Kirschbaum........90–4
Loop Systems Using Curvilinear Squares to Predict Root Location, Synthesis of Closed. Russell, Weaver........95–104
Losses in a D-C Motor When Operated from Grid-Controlled Rectifiers, Increased. Reiter, Ammerman........77–82

M

Machine Inputs, A Dial-Reading Translator for Digital. Klein, Gow....137–44; disc. 144
Machining, Theory of Electric Spark. Williams....105–08; disc 108
Magnetic Field at Altitude, Arc Interruption Phenomena in a. Dallas....419–22
Maintenance and Operating Experiences—Diesel-Electric Locomotives. Griffith, Pfeil....134–6; disc. 136
Measurement of Cathodic Protection Currents from Sacrificial Anodes. Hayward, Wainwright..350–2
Measurements, Aircraft Radio Interference. Newman, Schwantes, Stahmann....211–15
Mechanisms, Electrostatic Oscillation in Sensitive Contact. Stuelpnagel....397–401
Memory, Lead Networks Utilizing a Saturable Core. Scorgie....297–302
Method, Feedback Control Systems with Dead-Time Lag or Distributed Lag by Root-Locus. Chu....291–6
Methods for Optimum Servomechanism Design, Analogue. Fickeisen, Stout....244–50; disc. 250
Metropolitan Communities, Co-ordinated Transportation for Large. Kearns....54–7; disc. 57
Modulated Reversing Clutch Servo Systems, Consideration of Off-On-. Stuelpnagel, Dallas....406–10
Motor for Trolley Coaches, New Series Type. Woods, Steen....59–62
Motor Starting Controllers for Aircraft, Reduced Voltage D-C. Finison, Andersen....126–33
Motor When Operated from Grid-Controlled Rectifiers, Increased Losses in a D-C. Reiter, Ammerman....77–82
Motors, Stopping Devices for 400-Cycle. Zahorsky....185–90; disc. 190
Motors to Household Refrigeration Compressors, Application of. Packer....70–6; disc. 77
Motors, Voltage Rating Versus Horsepower of Synchronous and Induction. Miller....306–11

N

Network Analysis, A Block-Diagram Approach to. Stout....255–60
Networks Utilizing a Saturable Core Memory, Lead. Scorgie....297–302
Nonlinear Elements, Sinusoidal Analysis of Feedback-Control Systems Containing. Johnson....169–181
Nonlinear Feedback to Improve the Transient Response of a Servomechanism, The Use of. Lewis....449–53; disc. 453

O

One-Dimensional Closed-Loop Continuous Pursuit Task, An Electronic Apparatus for the Study of the Human Operator in a. Warren, Fitts, Clark. 19–23
Operating Experiences—Diesel-Electric Locomotives, Maintenance and. Griffith, Pfeil....134–6; disc. 136
Operator in a 1-Dimensional Closed-Loop Continuous Pursuit Task, An Electronic Apparatus for the Study of the Human. Warren, Fitts, Clark..19–23
Optimum Servomechanism Design, Analogue Methods for. Fickeisen, Stout....244–50; disc. 250
Oscillation in Sensitive Contact Mechanisms, Electrostatic. Stuelpnagel....397–401
Oscillation, Stabilization of a Servomechanism Subject to Large Amplitude. Sherrard....312–24
Overcurrent Protective Device Settings for Industrial Plants, Selecting A-C. Brightman....203–11

P

Pennsylvania Railroad, A New Electric Locomotive for the. Gowans, Widell, Bredenberg.27–35; disc. 35
Pennsylvania Railroad Ignitron-Rectifier Locomotive, The. Whittaker, Hutchison....37–42; disc. 42
Phase-Angle Loci, Synthesis of Feedback Control System by. Chu....330–9
Phenomena in a Magnetic Field at Altitude, Arc Interruption. Dallas....419–22
Philadelphia—Earth as a Heat Source, Residential Heat Pump Experiments in. Kiddler, Neher....343–50
Philadelphia—Installation and Operating Experience, Residential Heat Pump Experiments in. Harlow, Klapper....366–75
Philadelphia—Suggested Possibilities for Practical Applications, Residential Heat Pump Experiments in. Bary....361–5
Piecewise Linear Servomechanisms. Schwartz..401–05
Plant, Electric Power Supply for a Large Chemical. Friel, Smith....83–9
Plants, Selecting A-C Overcurrent Protective Device Settings for Industrial. Brightman....203–11
Power Contracts, Railway. Martin....24–6
Power Supply for a Large Chemical Plant, Electric. Friel, Smith....83–9
Power Systems, Unit Equipments for Aircraft Electric. Finison....145–9
Preheating of Electrolytic Tin Plate for Flow Brightening with High-Frequency Rotating Equipment, Induction. Thomas....13–18
Progress in Applications of Electrical Engineering to Air Transportation. Fritz....261–9
Protection Currents from Sacrificial Anodes, Measurement of Cathodic. Hayward, Wainwright...350–2
Protection for D-C Aircraft Generators, Feeder-Fault. Davis....430–5
Protection of the Bottoms of Inactivated Ships, The Use of Graphite Anodes for Cathodic. Oliver....324–9
Protective Device Settings for Industrial Plants, Selecting A-C Overcurrent. Brightman....203–11
Protective Scheme, A New Wheel-Slip. Smith....51–3
Pulsed Servos, An Approximate Transfer Function for the Analysis and Design of. Brown, Murphy....435–40; disc. 440
Pump Experiments in Philadelphia—Earth as a Heat Source, Residential Heat. Kidder, Neher..343–50
Pursuit Task, An Electronic Apparatus for the Study of the Human Operator in a 1-Dimensional Closed-Loop Continuous. Warren, Fitts, Clark....19–23

R

Radio Interference Measurements, Aircraft. Newman, Schwantes, Stahmann....211–15
Railroad, A New Electric Locomotive for the Pennsylvania. Gowans, Widell, Bredenberg....27–35; disc. 35
Railroad Ignitron-Rectifier Locomotive, The Pennsylvania. Whittaker, Hutchison..37–42; disc. 42.
Railway Electrification, Diesel-Electric Locomotives, and Some Future Aspects of Electric Traction. Huldschiner....48–50
Railway Power Contracts. Martin....24–6
Railway System, Electrification of Holland's. Candee....235–43; disc. 243
Reading Translator for Digital Machine Inputs, A Dial-, Klein, Gow....137–44; disc. 144
Rectifier Locomotive, The Pennsylvania Railroad Ignitron. Whittaker, Hutchison....37–42 disc. 42
Rectifiers, Increased Losses in a D-C Motor When Operated from Grid-Controlled. Reiter, Ammerman....77–82
Reduced Voltage D-C Motor Starting Controllers for Aircraft. Finison, Andersen....126–33
Refrigeration Compressors, Application of Motors to Household. Packer....70–6; disc. 77
Reliability of 28-Volt D-C Aircraft Electric Systems, Distribution System. Carlson, Sherrard...113–17
Requirements for Evaluating Blast-Cooled Aircraft Generators, Altitude Chamber. Scott....427–30; disc. 430
Residences Using the Heat Pump for Year-Round Air Conditioning, Load Characteristics of Five All-Electric. Sporn, Ambrose....353–60
Residential Heat Pump Experiments in Philadelphia—Earth as a Heat Source. Kidder, Neher...343–50
Residential Heat Pump Experiments in Philadelphia—Installation and Operating Experience. Harlow, Klapper....366–75
Residential Heat Pump Experiments in Philadelphia—Suggested Possibilities for Practical Applications. Bary....361–5
Response of a Servomechanism, The Use of Nonlinear Feedback to Improve the Transient. Lewis....449–53; disc. 453
Response of Airplanes by Means of Electric Equipment, Improving the Dynamic. Monroe....441–7; disc. 447
Reversing Clutch Servo Systems, Consideration of Off-On-Modulated. Stuelpnagel, Dallas....406–10
Review of the Factors Affecting the Temperature of Airborne Electric and Electronic Equipment, A. Lyons....423–5; disc. 425
Rise in Bundled Cables for Aircraft, Continuous Current and Temperature. Schach, Kidwell....376–84
Root Location, Synthesis of Closed Loop Systems Using Curvilinear Squares to Predict. Russell, Weaver....95–104
Root-Locus Method, Feedback Control Systems with Dead-Time Lag or Distributed Lag by. Chu....291–6
Rotating Electric Equipment, Study of Aircraft Cooling Systems for. Martin....150–60
Rotating Equipment, Induction Preheating of Electrolytic Tin Plate for Flow Brightening with High-Frequency. Thomas....13–18
Rotor Speed on the Weight of D-C Aircraft Generators, Effects of Terminal Voltage, Load Current and Minimum. Scott....191–7

S

Sacrificial Anodes, Measurement of Cathodic Protection Currents from. Hayward, Wainwright....350–2
Sampled-Data Systems, The Analysis of. Ragazzini, Zadeh....224–32; disc. 232
Saturable Core Memory, Lead Networks Utilizing a. Scorgie....297–302
Scheme, A New Wheel-Slip Protective. Smith....51–3
Selecting A-C Overcurrent Protective Device Settings for Industrial Plants. Brightman....203–11
Series Type Motor for Trolley Coaches, New. Woods, Steen....59–62
Servo Systems, Consideration of Off-On-Modulated Reversing Clutch. Stuelpnagel, Dallas....406–10
Servomechanism Design, Analogue Methods for Optimum. Fickeisen, Stout....244–50; disc. 250
Servomechanism Subject to Large Amplitude Oscillation, Stabilization of a. Sherrard....312–24
Servomechanism, The Use of Nonlinear Feedback to Improve the Transient Response of a. Lewis....449–53; disc. 453
Servomechanisms, Piecewise Linear. Schwartz.401–05
Servomechanisms, Stability Limits for Third-Order. Higgins, Levinthal....459–66; disc. 466
Servomechanisms, Stabilization Templates for. Smith....220–4
Servos, An Approximate Transfer Function for the Analysis and Design of Pulsed. Brown, Murphy....435–40; disc. 440
Settings for Industrial Plants, Selecting A-C Overcurrent Protective Device. Brightman....203–11
Ships, The Use of Graphite Anodes for Cathodic Protection of the Bottoms of Inactivated. Oliver....324–9
Short-Circuiting Techniques to Deactivate a Generator After a Fault. Markowitz, Pobst....182–4
Signal Component Control. Gimpel, Calvert...339–43
Sinusoidal Analysis of Feedback-Control Systems Containing Nonlinear Elements. Johnson....169–181
Slip Protective Scheme, A New Wheel-. Smith...51–3
Spark Machining, Theory of Electric. Williams....105–08; disc. 108
Squares to Predict Root Location, Synthesis of Closed Loop Systems Using Curvilinear. Russell, Weaver....95–104
Stability Limits for Third-Order Servomechanisms. Higgins, Levinthal....459–66; disc. 466
Stabilization of a Servomechanism Subject to Large Amplitude Oscillation. Sherrard....312–24
Stabilization Templates for Servomechanisms. Smith....220–4
Stabilizing Elements in Aircraft D-C Systems. Scorgie, Schaefer....387–96; disc. 396
Stopping Devices for 400-Cycle Motors. Zahorsky....185–90; disc. 190
Study of Aircraft Cooling Systems for Rotating Electric Equipment. Martin....150–60
Switchers, 15 Years' Progress in the Design of Industrial Diesel-Electric. Barrell....215–19
Synchronizing for Aircraft Alternators, Automatic. McCarty, Hills....454–7; disc. 458
Synchronous and Induction Motors, Voltage Rating Versus Horsepower of. Miller....306–11
Synthesis of Closed Loop Systems Using Curvilinear Squares to Predict Root Location. Russell, Weaver....95–104
Synthesis of Feedback Control System by Phase-Angle Loci. Chu....330–9
System Reliability of 28-Volt D-C Aircraft Electric Systems, Distribution. Carlson, Sherrard..113–17
System Transfer Functions by Inspection, The Writing of Closed-Loop Control. Flannigan, Kirschbaum....90–4
Systems, Characteristics of Faults on Aircraft Electric. Larson....165–9
Systems Containing Nonlinear Elements, Sinusoidal Analysis of Feedback-Control. Johnson...169–181
Systems for Rotating Electric Equipment, Study of Aircraft Cooling. Martin....150–60
Systems, Intermittent Faults in Aircraft Electric. Larson, Trbovich, Toomire....160–5
Systems, Unit Equipments for Aircraft Electric Power. Finison....145–9
Systems Using Curvilinear Squares to Predict Root Location, Synthesis of Closed Loop. Russell, Weaver....95–104

T

Techniques Applied to a Hydraulic Amplifier, Vacuum-Tube. Baring....302–06
Techniques to Deactivate a Generator After a Fault, Short-Circuiting. Markowitz, Pobst....182–4
Temperature of Airborne Electric and Electronic Equipment, A Review of the Factors Affecting the. Lyons....423–5; disc. 425
Temperature Rise in Aircraft Cables, Continuous Current and. Schach....197–203
Temperature Rise in Bundled Cables for Aircraft, Continuous Current and. Schach, Kidwell....376–84
Templates for Servomechanisms, Stabilization. Smith....220–4
Terminal Voltage, Load Current and Minimum Rotor Speed on the Weight of D-C Aircraft Generators, Effects of. Scott....191–7

Tests of Electrostatic Controls for Hazardous Industrial Applications. Beach..............1–12; disc. 12
Theory of Electric Spark Machining. Williams..............105–08; disc. 108
Tin Plate for Flow Brightening with High-Frequency Rotating Equipment, Induction Preheating of Electrolytic. Thomas..............13–18
Traction, Railway Electrification, Diesel-Electric Locomotives, and Some Future Aspects of Electric. Huldschiner..............48–50
Transfer Function for the Analysis and Design of Pulsed Servos, An Approximate. Brown, Murphy..............435–40; disc. 440
Transfer Functions by Inspection, The Writing of Closed-Loop Control System. Flannigan, Kirschbaum..............90–4
Transient Response of a Servomechanism, The Use of Nonlinear Feedback to Improve the. Lewis..............449–53; disc. 453
Translator for Digital Machine Inputs, A Dial-Reading. Klein, Gow..............137–44; disc. 144
(Transportation, Air) A Dial-Reading Translator for Digital Machine Inputs. Klein, Gow..............137–44; disc. 144
(Transportation, Air) A New Circuit Breaker for Aircraft Electric System. Beall, Reifschneider. 109–12
(Transportation, Air) A Review of the Factors Affecting the Temperature of Airborne Electric and Electronic Equipment. Lyons..............423–5; disc. 425
(Transportation, Air) Aircraft Radio Interference Measurements. Newman, Schwantes, Stahmann..............211–15
(Transportation, Air) Altitude Chamber Requirements for Evaluating Blast-Cooled Aircraft Generators. Scott..............427–30; disc. 430
(Transportation, Air) Automatic Synchronizing for Aircraft Alternators. McCarty, Hills..............454–7; disc. 458
(Transportation, Air) Characteristics of Faults on Aircraft Electric Systems. Larson..............165–9
(Transportation, Air) Continuous Current and Temperature Rise in Aircraft Cables. Schach.....197–203
(Transportation, Air) Continuous Current and Temperature Rise in Bundled Cables for Aircraft. Schach, Kidwell..............376–84
(Transportation, Air) Distribution System Reliability of 28-Volt D-C Aircraft Electric Systems. Carlson, Sherrard..............113–17
(Transportation, Air) Effects of Terminal Voltage, Load Current and Minimum Rotor Speed on the Weight of D-C Aircraft Generators. Scott..............191–7
(Transportation, Air) Electrostatic Oscillation in Sensitive Contact Mechanisms. Stuelpnagel...397–401
(Transportation, Air) Feeder-Fault Protection for D-C Aircraft Generators. Davis..............430–5
(Transportation, Air) Impedance Data for 400-Cycle Aircraft Distribution Systems. Exner, Singer..............410–19
(Transportation, Air) Improving the Dynamic Response of Airplanes by Means of Electric Equipment. Monroe..............441–7; disc. 447
(Transportation, Air) Intermittent Faults in Aircraft Electric Systems. Larson, Trbovich, Toomire..............160–5
(Transportation, Air) Reduced Voltage D-C Motor Starting Controllers for Aircraft. Finison, Andersen..............126–33
(Transportation, Air) Short-Circuiting Techniques to Deactivate a Generator After a Fault. Markowitz, Pobst..............182–4
(Transportation, Air) Stabilizing Elements in Aircraft D-C Systems. Scorgie, Schaefer. 387–96; disc. 396
(Transportation, Air) Stopping Devices for 400-Cycle Motors. Zahorsky..............185–90; disc. 190
(Transportation, Air) Study of Aircraft Cooling Systems for Rotating Electric Equipment. Martin..150–60
(Transportation, Air) Unit Equipments for Aircraft Electric Power Systems. Finison..............145–9
(Transportation, Air) Weight Analysis of Aircraft Actuators. Geyer, Treseder..............118–26
Transportation for Large Metropolitan Communities, Co-ordinated. Kearns..............54–7; disc. 57
(Transportation, Land) A New Control for Trolley Coaches. Willby..............63–5; disc. 65
(Transportation, Land) A New Electric Locomotive for the Pennsylvania Railroad. Gowans, Widell, Bredenberg..............27–35; disc. 35
(Transportation, Land) A New Wheel-Slip Protective Scheme. Smith..............51–3
(Transportation, Land) 15 Years' Progress in the Design of Industrial Diesel-Electric Switchers. Barrell..............215–19
(Transportation, Land) Maintenance and Operating Experiences—Diesel-Electric Locomotives. Griffith, Pfeil..............134–6; disc. 136
(Transportation, Land) New Series Type Motor for Trolley Coaches. Woods, Steen..............59–62
(Transportation, Land) Railway Electrification, Diesel-Electric Locomotives, and Some Future Aspects of Electric Traction. Huldschiner..............48–50
(Transportation, Land) Railway Power Contracts. Martin..............24–6
(Transportation, Land) The Pennsylvania Railroad Ignitron Rectifier Locomotive. Whittaker, Hutchison..............37–42; disc. 42
Transportation, Progress in Applications of Electrical Engineering to Air. Fritz..............261–9
Trolley Coaches, A New Control for. Willby..............63–5; disc. 65
Trolley Coaches, New Series Type Motor for. Woods, Steen..............59–62
Tube Techniques Applied to a Hydraulic Amplifier, Vacuum-. Baring..............302–06
28-Volt D-C Aircraft Electric Systems, Distribution System Reliability of. Carlson, Sherrard...113–17

U

Unit Equipments for Aircraft Electric Power Systems. Finison..............145–9

V

Vacuum-Tube Techniques Applied to a Hydraulic Amplifier. Baring..............302–06
Volt D-C Aircraft Electric Systems, Distribution System Reliability of 28-. Carlson, Sherrard......113–17
Voltage D-C Motor Starting Controllers for Aircraft, Reduced. Finison, Andersen..............126–33
Voltage, Load Current and Minimum Rotor Speed on the Weight of D-C Aircraft Generators, Effects of Terminal. Scott..............191–7
Voltage Rating Versus Horsepower of Synchronous and Induction Motors. Miller..............306–11

W

Weight Analysis of Aircraft Actuators. Geyer, Treseder..............118–26
Welding, Electric Glass. Shaw..............384–7
Wheel-Slip Protective Scheme, A New. Smith...51–3
Writing of Closed-Loop Control System Transfer Functions by Inspection, The. Flannigan, Kirschbaum..............90–4

2. Author Index

Ambrose, E. R.; P. Sporn. Load Characteristics of Five All-Electric Residences Using the Heat Pump for Year-Round Air Conditioning. (52-339)..............353–60
Ammerman, Charles R.; C. R. Reiter. Increased Losses in a D-C Motor When Operated from Grid-Controlled Rectifiers. (52-74)..............77–82
Andersen, W. C.; H. J. Finison. Reduced Voltage D-C Motor Starting Controllers for Aircraft. (52-134)..............126–33

B

Baring, J. A. Vacuum-Tube Techniques Applied to a Hydraulic Amplifier. (52-298)..............302–06
Barrell, P. W. Fifteen Years' Progress in the Design of Industrial Diesel-Electric Switchers. (52-169)..............215–19
Bary, Constantine. Residential Heat Pump Experiments in Philadelphia—Suggested Possibilities for Practical Applications. (52-340)..............361–5
Beach, Robin. Tests of Electrostatic Controls for Hazardous Industrial Applications. (52-1)...1–12
Beall, B. S., III; P. J. Reifschneider. A New Circuit Breaker for Aircraft Electric System. (52-131)..............109–12
Belsley, Steven E.; W. M. Kaufman. Disc.......448
Brecht, W. A. Disc..............35
Bredenberg, A.; F. D. Gowans, B. A. Widell. A New Electric Locomotive for the Pennsylvania Railroad. (52-23)..............27–35
Brightman, F. P. Selecting A-C Overcurrent Protective Device Settings for Industrial Plants. (52-184)..............203–11
Brown, B. M. Disc..............233
Brown, R. G.; G. J. Murphy. An Approximate Transfer Function for the Analysis and Design of Pulsed Servos. (52-314)..............435–40
Bruns, Robert A. Analogue Computers for Feedback Control Systems. (52-247)..............250–4
Burke, Harry E. Disc..............144

C

Calvert, J. F.; D. J. Gimpel. Signal Component Control. (52-211)..............339–43
Candee, A. H. Electrification of Holland's Railway System. (52-172)..............235–43; disc. 243
Carlson, K. W.; E. S. Sherrard. Distribution System Reliability of 28-Volt D-C Aircraft Electric Systems. (52-132)..............113–17
Carswell, Charles. Disc..............12
Chu, Yaohan. Feedback Control Systems with Dead-Time Lag or Distributed Lag by Root-Locus Method. (52-200)..............291–6
Chu, Yaohan. Synthesis of Feedback Control System by Phase-Angle Loci. (52-297)..............330–9
Clark, J. R.; C. E. Warren, P. M. Fitts. An Electronic Apparatus for the Study of the Human Operator in a 1-Dimensional Closed-Loop Continuous Pursuit Task. (52-8)..............19–23
Clauser, M. U.; J. G. Truxal. Disc..............447
Corfield, R. J. Disc..............136

D

Dallas, J. P. Arc Interruption Phenomena in a Magnetic Field at Altitude. (52-333)..............419–22
Dallas, J. P.; T. R. Stuelpnagel. Consideration of Off-On-Modulated Reversing Clutch Servo Systems. (52-329)..............406–10
Davis, William W. Feeder-Fault Protection for D-C Aircraft Generators. (52-332)..............430–5

E

Eisengrein, R. H. Disc..............458
Exner, D. W.; G. H. Singer, Jr. Impedance Data for 400-Cycle Aircraft Distribution Systems. (52-322)..............410–19

F

Fickeisen, F. C.; T. M. Stout. Analogue Methods for Optimum Servomechanism Design. (52-246)..............244–50
Finison, H. J. Unit Equipments for Aircraft Electric Power Systems. (52-135)..............145–9
Finison, H. J.; W. C. Andersen. Reduced Voltage D-C Motor Starting Controllers for Aircraft. (52-134)..............126–33
Fitts, P. M.; C. E. Warren, J. R. Clark. An Electronic Apparatus for the Study of the Human Operator in a 1-Dimensional Closed-Loop Continuous Pursuit Task. (52-8)..............19–23
Flannigan, J. B.; H. S. Kirschbaum. The Writing of Closed-Loop Control System Transfer Functions by Inspection. (52-81)..............90–4
Franklin, Paul W. Disc..............426
Friedman, Daniel. Disc..............425
Friel, A. C.; J. P. Smith. Electric Power Supply for a Large Chemical Plant. (52-80)..............83–9
Fritz, D. E. Progress in Applications of Electrical Engineering to Air Transportation. (52-266)..............261–9

G

Gellert, Max E. Disc..............447
Geyer, Howard M.; R. C. Treseder. Weight Analysis of Aircraft Actuators. (52-133)..............118–26
Gimpel, D. J.; J. F. Calvert. Signal Component Control. (52-211)..............339–43
Gow, Kenneth P.; A. L. Klein. A Dial-Reading Translator for Digital Machine Inputs. (52-118)..............137–44; disc. 144
Gowans, F. D.; B. A. Widell, A. Bredenberg. A New Electric Locomotive for the Pennsylvania Railroad. (52-23)..............27–35
Griffith, C. H.; E. H. Pfeil. Maintenance and Operating Experiences—Diesel-Electric Locomotives. (52-145)..............134–6

H

Harlow, J. H.; G. E. Klapper. Residential Heat Pump Experiments in Philadelphia—Installation and Operating Experience. (52-342)......366–75
Hatch, P. H. Disc..............43
Hayward, H. N.; R. M. Wainwright. Measurement of Cathodic Protection Currents from Sacrificial Anodes. (52-305)..............350–2
Higgins, T. J.; J. G. Levinthal. Stability Limits for Third-Order Servomechanisms. (52-313)..............459–66; disc. 466
Hills, G. W.; F. B. McCarty. Automatic Synchronizing for Aircraft Alternators. (52-311)......454–7
Huldschiner, G. Railway Electrification, Diesel-Electric Locomotives and Some Future Aspects of Electric Traction. (52-25)..............48–50
Hutchison, W. M.; C. C. Whittaker. The Pennsylvania Railroad Ignitron-Rectifier Locomotive. (52-24)..............37–42; disc. 46

J

James, H. B. Disc.........396
Johnson, E. Calvin. Sinusoidal Analysis of Feedback-Control Systems Containing Nonlinear Elements. (52-154).........169–181
Jones, Ray D. Disc.........190

K

Kauffman, William M.; S. E. Belsley. Disc.........448
Kearns, E. E. Co-ordinated Transportation for Large Metropolitan Communities. (52-36).........54–7; disc. 57
Kearns, E. E. Disc.........65
Kidder, A. H.; J. H. Neher. Residential Heat Pump Experiments in Philadelphia—Earth as a Heat Source. (52-341).........343–50
Kidwell, R. E., Jr.; M. Schach. Continuous Current and Temperature Rise in Bundled Cables for Aircraft. (52-323).........376–84
Kirschbaum, H. S.; J. B. Flannigan. The Writing of Closed-Loop Control System Transfer Functions by Inspection. (52-81).........90–4
Klapper, G. E.; J. H. Harlow. Residential Heat Pump Experiments in Philadelphia—Installation and Operating Experience. (52-342).........366–75
Klein, Arthur L.; K. P. Gow. A Dial-Reading Translator for Digital Machine Inputs. (52-118).........137–44; disc. 144
Koenig, J. F. Disc.........466

L

Larson, Lee R. Characteristics of Faults on Aircraft Electric Systems. (52-138).........165–9
Larson, L. R.; M. Trbovich, P. E. Toomire. Intermittent Faults in Aircraft Electric Systems. (52-137).........160–5
Levinthal, J. G.; T. J. Higgins. Stability Limits for Third-Order Servomechanisms. (52-313).........459–66; disc. 466
Lewis, J. B. The Use of Nonlinear Feedback to Improve the Transient Response of a Servomechanism. (52-199).........449–53; disc. 453
Linvill, William K. Disc.........234
Lyons, Leonard J. A Review of the Factors Affecting the Temperature of Airborne Electric and Electronic Equipment. (52-335)....423–5; disc. 426

M

MacLeod, D. R. Disc.........36, 42, 243
Markowitz, Oscar; J. W. Pobst. Short-Circuiting Techniques to Deactivate a Generator After a Fault. (52-139).........182–4
Martin, Cecil G. Study of Aircraft Cooling Systems for Rotating Electric Equipment. (52-136).150–60
Martin, T. M. C. Railway Power Contracts. (52-18).........24–6
McCarty, Fred B.; G. W. Hills. Automatic Synchronizing for Aircraft Alternators. (52-311).........454–7
McCloy, G. S. Disc.........76
McGee, P. A. Disc.........44
Miller, C. E. Voltage Rating Versus Horsepower of Synchronous and Induction Motors. (52-308).........305–11
Monroe, William R. Improving the Dynamic Response of Airplanes by Means of Electric Equipment. (52-338).........441–7; disc. 448
Moore, H. G. Disc.........65
Mueller, E. L. Disc.........65
Murphy, G. J.; R. G. Brown. An Approximate Transfer Function for the Analysis and Design of Pulsed Servos. (52-314).........435–40

N

Neher, J. H.; A. H. Kidder. Residential Heat Pump Experiments in Philadelphia—Earth as a Heat Source. (52-341).........343–50
Newman, M. M.; R. C. Schwantes, J. R. Stahmann. Aircraft Radio Interference Measurements. (52-241).........211–15
Nims, Paul T. Disc.........250

O

Oetting, R. L. Electric Lighting in the First Century of Engineering. (52-267).........269–91
Ogden, H. S. Disc.........46
Oliver, J. P. The Use of Graphite Anodes for Cathodic Protection of the Bottoms of Inactivated Ships. (52-306).........324–9

P

Packer, L. C. Application of Motors to Household Refrigeration Compressors. (52-69).........70–6; disc. 77
Perkinson, T. F. Disc.........243
Pfeil, E. H.; C. H. Griffith. Maintenance and Operating Experiences—Diesel-Electric Locomotives. (52-145).........134–6
Pines, S. Disc.........447
Pobst, J. W.; O. Markowitz. Short-Circuiting Techniques to Deactivate a Generator After a Fault. (52-139).........182–4

R

Ragazzini, J. R.; L. A. Zadeh. The Analysis of Sampled-Data Systems. (52-161).........225–32; disc. 234
Ragazzini, John R.; L. A. Zadeh. Disc.........440
Rea, J. B. Disc.........448
Reifschneider, P. J.; B. S. Beall, III. A New Circuit Breaker for Aircraft Electric System. (52-131).........109–12
Reiter, C. Richard; C. R. Ammerman. Increased Losses in a D-C Motor When Operated from Grid-Controlled Rectifiers. (52-74).........77–82
Russell, D. W.; C. H. Weaver. Synthesis of Closed Loop Systems Using Curvilinear Squares to Predict Root Location. (52-82).........95–104
Rutan, E. J. Disc.........12

S

Salzer, John M. Disc.........232
Schach, M. Continuous Current and Temperature Rise in Aircraft Cables. (52-141).........197–203
Schach, Milton; R. E. Kidwell, Jr. Continuous Current and Temperature Rise in Bundled Cables for Aircraft. (52-323).........376–84
Schaefer, D. H.; D. G. Scorgie. Stabilizing Elements in Aircraft D-C Systems. (52-330). 387–96; disc. 397
Schwantes, R. C.; M. M. Newman, J. R. Stahmann. Aircraft Radio Interference Measurements. (52-241).........211–15
Schwartz, James W. Piecewise Linear Servomechanisms. (52-344).........401–05
Scorgie, D. G. Lead Networks Utilizing a Saturable Core Memory. (52-272).........297–302
Scorgie, D. G.; D. H. Schaefer. Stabilizing Elements in Aircraft D-C Systems. (52-330).........387–96; disc. 397
Scott, Dale H. Altitude Chamber Requirements for Evaluating Blast-Cooled Aircraft Generators. (52-336).........427–30; disc. 430
Scott, Dale H. Effects of Terminal Voltage, Load Current and Minimum Rotor Speed on the Weight of D-C Aircraft Generators. (52-142).........191–7
Shaw, M. R., Jr. Electric Glass Welding. (52-320).........384–7
Sherrard, Emile S. Stabilization of a Servomechanism Subject to Large Amplitude Oscillation. (52-277).........312–24
Sherrard, E. S.; K. W. Carlson. Distribution System Reliability of 28-Volt D-C Aircraft Electric Systems. (52-132).........113–17
Simonds, R. H. Disc.........430
Singer, G. H., Jr.; D. W. Exner. Impedance Data for 400-Cycle Aircraft Distribution Systems. (52-322).........410–19
Smith, Gerald F. Disc.........46
Smith, J. P.; A. C. Friel. Electric Power Supply for a Large Chemical Plant. (52-80).........83–9
Smith, Otto J. M. Stabilization Templates for Servomechanisms. (52-239).........220–4
Smith, R. M. A New Wheel-Slip Protective Scheme. (52-35).........51–3
Smith, S. V. Disc.........36, 44
Sporn, Philip; E. R. Ambrose. Load Characteristics of Five All Electric Residences Using the Heat Pump for Year-Round Air Conditioning. (52-339).........353–60
Stahmann, J. R.; M. M. Newman, R. C. Schwantes. Aircraft Radio Interference Measurements. (52-241).........211–15
Stallard, David, B. Disc.........453
Steen, C. R.; G. M. Woods. New Series Type Motor for Trolley Coaches. (52-37).........59–62
Stineman, Russel W. Disc.........458
Stout, T. M. A Block-Diagram Approach to Network Analysis. (52-254).........255–60
Stout, T. M.; F. C. Fickeisen. Analogue Methods for Optimum Servomechanism Design. (52-246).........244–50
Stuelpnagel, Thomas R. Electrostatic Oscillation in Sensitive Contact Mechanisms. (52-328)..397–401
Stuelpnagel, T. R.; J. P. Dallas. Consideration of Off-On-Modulated Reversing Clutch Servo Systems. (52-329).........406–10

T

Thomas, William T. Induction Preheating of Electrolytic Tin Plate for Flow Brightening with High-Frequency Rotating Equipment. (52-4).....13–18
Toomire, P. E.; L. R. Larson, M. Trbovich. Intermittent Faults in Aircraft Electric Systems. (52-137).........160–5
Trbovich, M.; L. R. Larson, P. E. Toomire. Intermittent Faults in Aircraft Electric Systems. (52-137).........160–5
Treseder, R. C.; H. M. Geyer. Weight Analysis of Aircraft Actuators. (52-133).........118–26
Truxal, J. G.; M. U. Clauser. Disc.........447

W

Wainwright, R. M.; H. N. Hayward. Measurement of Cathodic Protection Currents from Sacrificial Anodes. (52-305).........350–2
Warren, C. E.; P. M. Fitts, J. R. Clark. An Electronic Apparatus for the Study of the Human Operator in a 1-Dimensional Closed-Loop Continuous Pursuit Task. (52-8).........19–23
Weaver, C. H.; D. W. Russell. Synthesis of Closed Loop Systems Using Curvilinear Squares to Predict Root Location. (52-82).........95–104
Weber, I. Disc.........108
White, A. P. Disc.........76
Whittaker, C. C.; W. M. Hutchison. The Pennsylvania Railroad Ignitron-Rectifier Locomotive. (52-24).........37–42; disc. 46
Widell, B. A.; F. D. Gowans, A. Bredenberg. A New Electric Locomotive for the Pennsylvania Railroad. (52-23).........27–35
Willby, N. H. A New Control for Trolley Coaches. (52-38).........63–5; disc. 65
Williams, Everard M. Theory of Electric Spark Machining. (52-101).........105–08; disc. 108
Williamson, J. W. Bus Bar Design for High-Frequency Induction Heating. (52-68).........66–9
Woods, G. M. Disc.........57
Woods, G. M.; C. R. Steen. New Series Type Motor for Trolley Coaches. (52-37).........59–62

Z

Zadeh, L. A.; J. R. Ragazzini. The Analysis of Sampled-Data Systems. (52-161).........225–32; disc. 234
Zadeh, Lotfi; J. R. Ragazzini. Disc.........440
Zahorsky, L. A. Stopping Devices for 400-Cycle Motors. (52-140).........185–90; disc. 190

Preface

The AIEE *Transactions* for 1952 (volume 71) is published in three parts: Part I. Communication and Electronics; Part II. Applications and Industry; and Part III. Power Apparatus and Systems. The papers in each of the three parts are classified according to subject matter as follows:

Part I. Communication and Electronics

- Communication Switching Systems
- Radio Communications Systems
- Special Communications Applications
- Telegraph Systems
- Television and Aural Broadcasting Systems
- Wire Communications Systems
- Basic Sciences
- Computing Devices
- Electrical Techniques in Medicine and Biology
- Electronic Power Converters
- Electronics
- Instruments and Measurements
- Magnetic Amplifiers
- Metallic Rectifiers
- Nucleonics

Part II. Applications and Industry

- Air Transportation
- Domestic and Commercial Applications
- Land Transportation
- Marine Transportation
- Production and Application of Light
- Chemical, Electrochemical and Electrothermal Applications
- Electric Heating
- Electric Welding
- Feedback Control Systems
- General Industry Applications
- Industrial Control
- Industrial Power Systems
- Mining and Metal Industry

Part III. Power Apparatus and Systems

- Carrier Current
- Insulated Conductors
- Power Generation
- Protective Devices
- Relays
- Rotating Machinery
- Substations
- Switchgear
- System Engineering
- Transformers
- Transmission and Distribution

- Education
- Safety
- Research

Each part has been indexed separately in the back of that particular part. The three parts are not cross-referenced, hence the user should determine first whether the subject matter of the paper desired is in the field of communication and electronics, applications and industry, or power apparatus and systems. Papers are listed in the subject index under several key words in the titles. The original numbers assigned to the papers are given in the author index.

Volume 71 contains the technical papers and related discussions presented at these meetings:

1. Winter General Meeting, New York, N. Y., January 21–25, 1952.
2. South West District Meeting, St. Louis, Mo., April 15–17, 1952.
3. North Eastern District Meeting, Binghamton, N. Y., April 30–May 2, 1952.
4. Summer General Meeting, Minneapolis, Minn., June 23–27, 1952.
5. Pacific General Meeting, Phoenix, Ariz., August 19–22, 1952.
6. AIEE Participation in Centennial of Engineering, Chicago, Ill., September 10–12, 1952.
7. Fall General Meeting, New Orleans, La., October 13–17, 1952.
8. Middle Eastern District Meeting, Toledo, Ohio, October 28–30, 1952.

The papers presented at the first three meetings were issued as AIEE *Proceedings.* These papers have been correlated according to subject matter, and combined with the papers presented at the remaining meetings which were published in the three bimonthly publications inaugurated in July and August, comprise Parts I, II, and III of volume 71 of the 1952 *Transactions.* Ultimately each part of the *Transactions* will contain the same papers and discussions that will have been published in the six successive issues of the corresponding bimonthly publication.

Statements and opinions given in the papers and discussions published in *Transactions* are the expressions of the contributors for which the American Institute of Electrical Engineers assumes no responsibility.

Tests of Electrostatic Controls for Hazardous Industrial Applications

ROBIN BEACH
FELLOW AIEE

THE destruction and casualties from fires and explosions which originate from electrostatic ignition in hazardous processes are all too well known among safety engineers and industrial management to require elaboration. The annual property loss to industry is reported to be approximately $100,000,000 exclusive of costly losses from indirect sources. These losses, as well as the injuries and deaths of operators resulting from electrostatic origin, occur where flammable components are ignited in industrial processes. Industrial solvents and flammable or explosive chemicals are the chief offenders.

The ravages of static electricity occur quite generally throughout the gamut of industrial processes wherever the stock or products manufactured and used possess insulating properties. Static electricity variously injures the quality of products, wastes stock, slows down production, causes fire and explosion losses, and inflicts injuries on personnel. In a few typical operations among the textile, paper, rubber, and plastics industries, conventional means of controlling static electricity are applicable by the use of so-called static neutralizers.

In many hazardous industrial operations, on the other hand, no simple and ready-made antistatic expedients are available. The host of specialized problems here can be resolved only first, by making a detailed investigation into the sources of the generation and accumulation of static electricity, the physical properties of the stock, the magnitude of the electrostatic voltages, the intensity of spark energies, and the properties of mixtures of flammable components and second, by applying tailor-made prescriptions for the control of the static electricity causing potential fire and explosion hazards.

Only a small minority of industrial personnel, those who happen to be safety conscious and, therefore, who wisely anticipate dangerous conditions within their plants, take specialized counsel on applying timely remedial steps for instituting appropriate control before costly disasters drive home their forceful lessons.

As stated, the manufacturing and processing of textiles, paper, plastics, and rubber are, by and large, the only areas of industry in which ready-made devices for the control of static electricity directly apply.[1] Quite generally, safety engineers or others who are commissioned in industrial plants to institute safe practices are expected to select from the market those static neutralizers which they hope will satisfactorily resolve their particular electrostatic hazards. Unless these individuals are well acquainted with the basic principles and specific applicability of all the neutralizers on the market, which is generally not the situation, the attainment of the best resolution of their difficulties becomes speculative and uncertain.[2] To the author's knowledge no technical information is available showing test performance of static neutralizers which enables one to select antistatic equipment with a predetermined degree of assurance or satisfaction.

In order to assist safety engineers and plant personnel in a more discriminating choice of suitable antistatic equipment, the General Industry Applications Committee of the AIEE requested the author to prepare a comprehensive paper classifying the various types of static neutralizers, explaining their relative advantages and limitations, appraising their degrees of safety, and presenting test data to show the comparative effectiveness of static neutralizers within each classification.

The author has no thought of representing himself as judge and jury in attempting to adjudicate the relative merits of one type of neutralizer over another. In this regard, the test data disclosing the discharge properties and other aspects of the various antistatic units clearly indicate their performance characteristics. From his years of experience in and research studies of industrial electrostatics, the author is taking the liberty of interpreting some of the disclosures from his test data which he believes will be helpful both to the manufacturers of neutralizer equipment and to prospective purchasers.

The Basic Principle of Discharge of Electrified Stock

All industrial static neutralizers utilize the principle of ionizing the air at locations where the offending electrification on the processed stock exists—whereby vast numbers of negative and positive charges are created in the intervening space between the units and the stock. The electric charges thus formed in the air which possess opposite polarity to those generated and accumulated on the stock are attracted to, and thus neutralize the charges on, the electrified stock. If a sufficient number of charges in the ambient ionized air are available to neutralize all of the charges of opposite polarity on the electrified stock, then complete discharge of the processed stock occurs. Except for a very few specially hazardous

---

Paper 52-1, recommended by the AIEE General Industry Applications Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted January 30, 1951; made available for printing November 9, 1951.

ROBIN BEACH is with Engineers Associated, Brooklyn, N. Y.

The author wishes to express his appreciation for the co-operation given him by the nine manufacturers of electrostatic equipment in providing their neutralizers and supplemental equipment for test. Also, he acknowledges his indebtedness for aid and encouragement to Col. E. F. McCrossin of McCrossin and Company, and to W. J. Hinson of Hinson-McAuliffe Corporation.

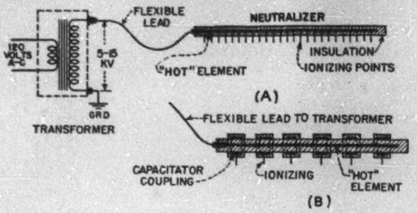


Figure 1(A). "Hot" electrode of an a-c neutralizer conductively connected to ionizing points and energized from the high potential terminal of the transformer. The electrified stock in passing near the ionizing points is discharged by attracting from the air ions of opposite polarity. **Figure 1(B).** The capacitor coupling of the ionizing points to the "hot" electrode which provides safety in preventing potential fatalities from electric shock

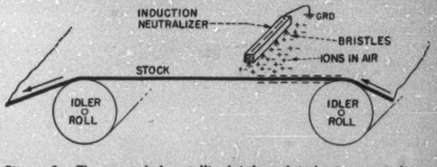


Figure 2. The grounded metallic bristles of induction neutralizers attain high potential gradients at their tips which originate by concentration of the electric field from the electrified stock, thus causing the air to be ionized and consequently the stock discharged

processes, complete discharge of the electrified stock is quite unnecessary. Where fire and explosion hazards exist, of course, safe practice prescribes the reduction of electrification to a value well below the spark-ignition threshold.

The ionization of air is a process whereby outer planetary electrons are ejected from otherwise neutral atoms and molecules of the gases and vapors which constitute the air. The products of ionization, consisting always of pairs of negative and positive ions, are generally electrons from the ionized atoms, as the negative charges, and the residual atoms or molecules each minus at least one electron, as the positive charges. The former are exceedingly mobile because of their tiny mass while the latter, being thousands of times heavier than electrons, are relatively slow in their movement through space.

High-Voltage Preionization by Neutralizers With Proximity Grounds

One method of causing ionization, referred to as preionization, employs high-voltage stressing of the air between electrified point electrodes and nearby grounded members; the latter for most effective ionization constituting a built-in part of the neutralizer unit and being herein called a proximity ground. The high potential impressed on the ionizing points is provided from one terminal of the high-voltage coil of a transformer while the other terminal of the coil is connected to the ground rod, as illustrated in Figure 1(A). When the electric stress at the points attains the breakdown value for air at a potential gradient in the order of 31 kv per centimeter, under standard atmospheric conditions, ionization of the air is initiated in the intervening electrified space between the points and the grounded member. If the applied electric stress is increased progressively above this threshold value, the ionization of the air greatly multiplies until final spark-over occurs.[3] The safe operation of neutralizers prescribes an ionizing intensity well below possible spark-over voltages and, especially so, where flammable vapor-air mixtures are present.

Obviously, ionizing points connected conductively to the hot element of the neutralizer and energized directly from a high-voltage transformer, as shown in Figure 1(A), introduces hazards of fatal shock to an individual who touches or receives sparks from the electrified points.[4] One means for reducing this potential hazard which is utilized by all high-voltage neutralizers comprises the use of capacitive coupling of the ionizing points to the hot element of the neutralizer bar as shown in Figure 1(B). This and other safety provisions are discussed later in the paper.

Since this method of creating ionization by the use of high-voltage excitation of points is not dependent on the presence or physical state of the processed stock, a neutralizer employing this principle is classified as a high-voltage preionizing unit with proximity ground.

The Physical Mechanism of High Voltage Ionization

When a voltage capable of creating ionization is impressed across, say, a one centimeter gap between two parallel plates of a conventional air-insulated capacitor, an electron at the negative plate, by force of repulsion, quickly acquires a velocity toward the positive plate in the order of 20,000,000 centimeters per second. Through its collisions en route with neutral atoms and molecules of the intervening air, their outermost planetary electrons are knocked out, when the energy of impact becomes high enough, thereby causing the formation of pairs of negative and positive ions. This process of creating negative and positive ions is referred to as ionization by collision. Actually, a high-velocity free electron gives rise to a tremendous and increasing multiplication of electrons and positive ions as it and its progeny of electrons advance toward and finally reach the positive plate. This characteristic growth of electron population is known as an electron avalanche.

In order to illustrate numerically the enormous growth of electron population during a single electron-initiated avalanche, assume the potential gradient at the negative plate of the condenser to be at about 31 kv per centimeter and also assume the availability of new initiating electrons at the negative plate at the rate of one each microsecond, created by virtue of ultraviolet irradiation directed from outside into the gap space and onto the negative plate. Under these assumptions, the threshold energy required for the initiating electron to create the first pair of ions by collision occurs about 0.04 centimeter from the negative electrode.[5] Halfway across the gap about 5,000 electron and positive ion pairs are formed. Three quarters of the way across the gap, nearly 500,000 electron and positive ion pairs are formed; and, at a distance of about 0.04 centimeter from the positive plate, the electron and positive ion population has increased to nearly 12,000,000. The electrons, for the most part, reach their destination at the positive plate during the electron avalanche, as described, in approximately 0.1 microsecond. The low mobility of the heavy positive ions, which is in the order of 0.01 to 0.0001 that of electrons, causes the positive ions during this short period to remain practically at their locations of birth in the process of ionization.

The replacement of the two parallel condenser plates, as cited in the above illustration, with a capacitor composed of an alignment of negatively electrified metallic points and a ground rod, as used in a high-voltage preionizing neutralizer with proximity ground, affects greatly the electric field distribution in the gap, changing it from one of uniform value as in the former to one of high distortion in the latter, especially near the points. The concentration of electric field at a needle point, when it is negatively electrified, greatly increases the potential gradient there for a given applied voltage, a condition which vastly multiplies the ionization activity at the negative point and accounts for the much lower spark-over voltages for needle-point gaps.

If the point of a point-to-plane capacitor system is electrified to positive polarity, the formation of negative and positive ions in the gap is markedly decreased in comparison to that for a negative point. Of course, the geometry of both electrodes plays an important part in the process and magnitude of ionization by collision. The ground rod in neutralizers, by being smooth and of extensive area compared with the point, lacks the high potential gradient which characterizes a point electrode; and, for this reason in part, the degree of ionization within the gap is decimated. When the point-to-ground-rod system of electrodes for neutralizers is energized with alternating current, as all high-voltage preionizing units are, copious ionization occurs for a part of the half-cycle when the point is highly electrified to negative polarity. The ionization, on the other hand, is relatively small for the succeeding half-cycle when the point is electrified to positive polarity.

When the electrodes are energized by alternating current, no ionization occurs at all during the low-voltage phase both of the positive and negative half cycles. In this interim period of 60 electrical degrees, more or less, only the process of recombination of electrons and positive ions takes place. Of course, the recombination of ions goes on continuously during ionization periods by which process photons of radiant energy are liberated. When these photon units of energy are liberated in sufficient numbers that at least five of them reach the retinal rod receptors of the dark-adapted eye during exposures of one millisecond, the threshold of visual sensation is attained and corona glow first becomes visible.[6]

The dissymmetrical behavior of the cyclical ionization from a-c excitation of various neutralizers may be observed from a cathode-ray oscilloscope. The current waves possess a phase-leading characteristic caused by the capacitive nature of the circuit. They are peculiarly distorted as a result of the variable conduction under deionization during the low-voltage periods when the ions undergo excessive recombination. Since the ionization is greatest when the voltage wave attains its maximum value, the maxima of the current waves occur about in phase with it. The higher current on the negative half-cycle of voltage clearly indicates the increased ionization from the points during negative polarity.

Several of the high-voltage neutralizers which utilize proximity grounds embody the ground feature in the form of an outer metal encasement rather than a ground rod. The ionizing points, in some of these, being the ends of flexible fine wire bristles, are insulated from the grounded encasement by vulcanite or ceramic centering elements.

High-Voltage Preionization by Neutralizers Without Proximity Grounds

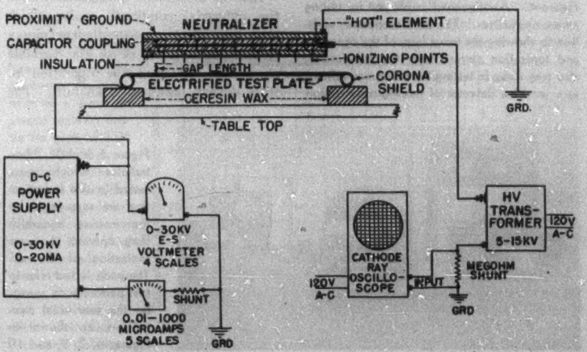
Neutralizer bars which do not employ proximity grounds also cause ionization of the ambient air when energized although much less copiously. A few static neutralizers are constructed to operate in this manner. Here, too, the discharge points, as a safety provision, are electrified through capacitive coupling to the hot element of the neutralizer bar. However, the intensity of the electric field between the ionizing points and the remote grounded parts of the processing machine, being greatly attenuated, causes a relatively low potential gradient at the electrified points; a condition which is reflected by a marked reduction of ionization. In order to increase the relatively low ionization at the points, transformer voltages in the order of 12 to 15 kv are utilized in these neutralizers whereas, in neutralizers employing built-in proximity grounds, transformer voltages in the order of 5 kv are more commonly employed.

Preionization by Radioactive Neutralizers

Another means for ionizing air utilizes the emission of alpha particles from the spontaneously disintegrating nuclei of radium, polonium, or other radioactive isotopes of the heavy metal elements. Alpha particles, being the emitted components from the nuclei of radioactive isotopes, have been found to consist each of a helium nucleus, polarized with two positive electric charges. The ionizing range of alpha emission in air is limited to an effective distance of about three inches. The ionizing process by which alpha emission liberates electrons from neutral atoms in the air comprises high-energy collision. Neutralizers which utilize alpha emission from radioactive isotopes to ionize air also are classed as preionizers.

A radioactive neutralizer which contains a radium activator possesses a theoretical advantage of unlimited life, since the half-life period of radium is 1,690 years. In practice, however, the effective life of radium as a neutralizer is determined by the length of time the exposed side of the thin gold and nickel envelope containing the activator can be cleaned of alpha-absorbing deposits without inflicting permanent injury to the fragile metallic membranes. A disadvantage of the radium neutralizer, aside from its eventual loss of effectiveness by absorption of alpha radiation, lies in its additional emission of beta and gamma rays.

Figure 3. Electric circuits for the d-c electrification of the test plate and for the a-c excitation of the neutralizer



The former possess distinctive erythemal properties which, in prolonged exposure, injure skin tissue while the latter, like hard X rays, penetrate deeply into body tissue and cause harmful secondary emission.[7] Hence, in their applications, radium neutralizers should be so located that operators are never subjected to prolonged periods of direct or reflected beta or gamma radiation.

Polonium, quite unlike radium, possesses the very short half-life period of 144 days, although neutralizers using polonium as an activator are claimed to have an effective life of nearly two years. An advantage possessed by polonium neutralizers results from its emission of alpha particles without appreciable amounts of the detrimental beta or gamma radiation. In addition to its economic disadvantage of short life, polonium is reported to possess toxic effects by possible ingestion of microscopic flakes resulting through contact or inhalation. Hence, the active element in the neutralizer is electroplated with a thin layer of gold as a means of providing protection against these hazards.[8]

Ionization by Electrostatic Induction Neutralizers

The second general classification of neutralizers is one which employs the principle of ionization which is dependent on electrostatic induction rather than preionization. Neutralizers of this class comprise a miscellany of grounded bars wrapped loosely with tinsel, grounded jewelers' wire, and various configurations of grounded metallic bristles and arrays of sharp points which are variously disposed along metallic or wooden supporting bars. In every instance, these neutralizers derive their ionizing facility by electrostatic induction from the electric field of the electrified stock. One such induction neutralizer is illustrated with relation to the stock it is discharging in Figure 2. The voltage between the electrified stock and the tips of the grounded metallic bristles electrically stresses the intervening air. As the potential gradient at the tips of the grounded bristles attains the threshold value for ionization, discharging action by the liberated electrons and positive ions is initiated at the surface of the electrified stock. When the potential gradient at the tips of the bristles exceeds the threshold of ionization, the discharging facility of the induction neutralizer greatly increases, as the ionization is an exponential function of the increased induction voltage.

Since both the initiation of ionization and the subsequent increase of ionization are dependent on the potential gradients at the tips of the bristles, the design of these neutralizers is critically influenced by the manner in which the bristles are constructed, grouped, and disposed along the neutralizer bar. The disadvantage of induction neutralizers lies in their inability to start discharging electrified stock until the potential gradient at the tips of the grounded bristles attains the threshold value for ionization. Because of this inherent characteristic, neutralizers in this class are obviously not suited for application to the neutralization of stock whose electrification is too low to induce threshold potential gradients at the tips of the grounded bristles. On the other hand, where stock is highly electrified, induction neutralizers when properly constructed possess definite physical and economic advantages, several of which are explained later in this paper.[9]

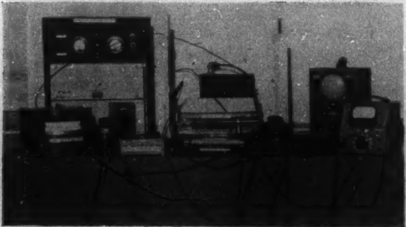


Figure 4. Arrangement employed for testing an a-c neutralizer. The oscilloscope, in addition to showing the wave form of the capacitor and ionization current from the transformer, also may serve in testing induction neutralizers as a sensitive detector of threshold ionization

Testing the Discharge Performance of Neutralizers

In order to simulate by laboratory tests the industrial conditions under which static neutralizers perform their service, a scrutinizing analysis of all the variable parameters affecting the operation of these devices becomes paramount. The essential element in a method to test the ionizing facility of static neutralizers comprises an appropriate substitute for the electrified stock. Unlike the electrified stock, the substitute element should be free from the disturbing factor of variable potential from the electrification and also free from the uncertain influences of spurious ionization, leakage, and corona from the stock.

An anticorona polished metallic plate was employed to represent the electrified stock. This plate was electrified from an appropriate d-c power supply with controls to apply any desired voltage up to 15 kv. The practical elimination of leakage current from the test circuit was achieved by insulating the test plate and other high-voltage parts from the work bench by blocks of ceresin wax, a substance which provides the highest known degree of insulation, as well as possessing a flat resistivity characteristic with respect to relative humidity up to a value of 96 per cent. The corona disturbances in the high-voltage circuitry were minimized by the use of adequate wire diameter, by substantial jackets of polyethylene insulation on the wire, and by the elimination of the ionizing effects from the sharp ends of wires through covering the wire tips with globules of solder.

The static neutralizer, under test, may then be suspended across the test plate and equipped with its supplementary facilities as provided for its industrial service. These facilities consist of energizing the static bar with its rated transformer voltage in preionizer units, of

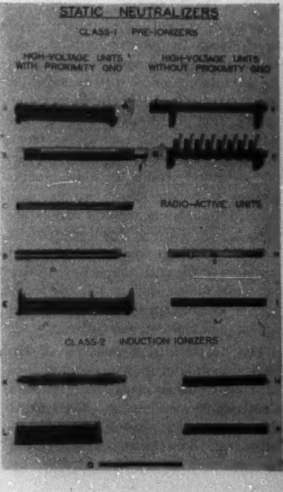

STATIC NEUTRALIZERS
CLASS-1 PRE-IONIZERS
HIGH-VOLTAGE UNITS WITH PROXIMITY GND
HIGH-VOLTAGE UNITS WITHOUT PROXIMITY GND
A
B
C
RADIO-ACTIVE UNITS
D
E
CLASS-2 INDUCTION IONIZERS
K
L
O

Figure 5 (right). Neutralizers which were tested in this investigation are segregated by the manner in which they operate to cause ionization of the air. The code letters refer to the performance curves for the individual neutralizers as shown in Figures 6, 7, 9, and 10

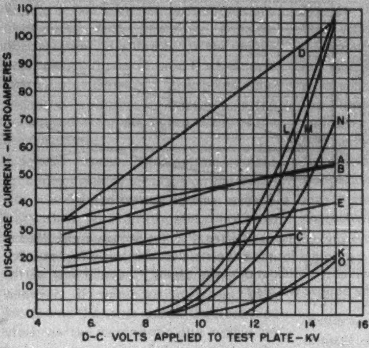


Figure 6 (left). Gap length= 1.00 inch

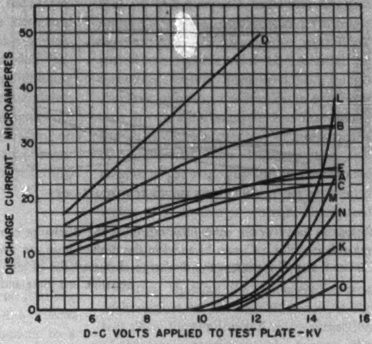


Figure 7 (right). Gap length= 1.50 inches

Curves A, B, C, D, and E represent the discharge characteristics of high-voltage preionizing neutralizers—each with a proximity ground. Curves K L, M, N, and O represent the discharge characteristics of induction neutralizers. The test plate was electrified to negative polarity in each of these tests

grounding the proximity ground, if such is used, or of grounding the housing of radioactive neutralizers or the metallic bristles of induction neutralizers. The gap spacing between the neutralizer bar and the test plate may be conveniently adjusted along the length of the bar to the desired value with suitable precision by the use of machined gauges. The gap lengths employed in the tests were 0.75, 1.00, 1.25, 1.50, and 2.00 inches.

When the neutralizers were operating under their prescribed conditions, their reactiveness in discharging the test plate was measured by the value of direct current supplied to the test plate from the d-c power supply at the impressed d-c voltage. The d-c voltage to which the test plate was electrified was adjusted throughout the tests to cardinal values ranging from zero to 15 kv and the voltage was measured by the use of a calibrated special electrostatic voltmeter. The discharge current was indicated on a calibrated D'Arsonval microammeter which could be read to 0.01 microampere. The test circuit is shown schematically in Figure 3; and the equipment utilized during the discharge tests of a static neutralizer is illustrated in Figure 4. The instrumental precision of the test results for a given gap setting could be duplicated quite generally within 2 per cent and for a resetting of the gap within about 5 per cent.

In practice, the discharge of highly electrified stock composed of fibrous materials, such as textiles and paper, is augmented by the fibers attracted from the surface into the gap not only by virtue of the ionizing properties of the fine semiconductive fibers but, also, by the reduction they cause in the effective length of the gap. In plastic and rubber stock, the smooth surfaces of these materials largely remove these variable ionizing factors except for slight effects from roughened edges. The anticorona polished test plate was intentionally designed to eliminate from the tests these variable factors which inherently contribute by spurious ionization to the discharge from electrified stock.

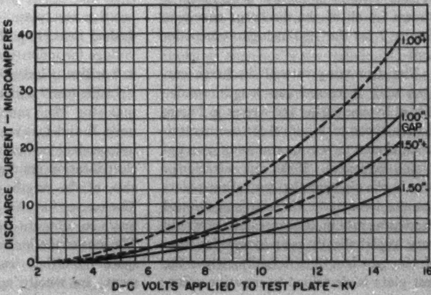
Also, the varying d-c voltage, existing between the electrified stock and ground, which the neutralizers are expected to reduce or eliminate, introduces into industrial applications another variable factor. In the discharge tests, this uncertainty was precluded by maintaining constant voltage at the electrified test plate. Hence, the discharge facility of each neutralizer during test was accurately represented by the direct current through the microammeter from the d-c power supply to the test plate.

Since the processing of stock over steel rolls most commonly electrifies the stock with negative polarity, the test plate was electrified to negative potential throughout the tests. However, in the test of every neutralizer, at least one set of data was taken for comparative purposes with the test plate charged to positive polarity.

The Discharge Facility of Neutralizers

The neutralizer units supplied by the manufacturers were all prepared to a 12 inch effective length. The units which were tested are displayed for reference in the reproduction shown by Figure 5. They are segregated into two general classifications, *class 1*–of preionizers and *class 2*–of induction ionizers. In the preionizing classification, the neutralizers may function from high-voltage excitation either with or without benefit of

Figure 8. The upper and lower curves in the solid lines represent the discharge characteristics of a 4-mil diameter metallic bristle for gaps of 1.00 and 1.50 inches, respectively, with the test plate electrified to negative polarity. The upper and lower curves in the dashed lines similarly represent the discharge curves of the bristles but with the test plate electrified to positive polarity



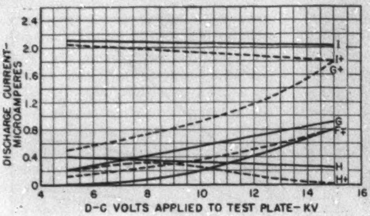
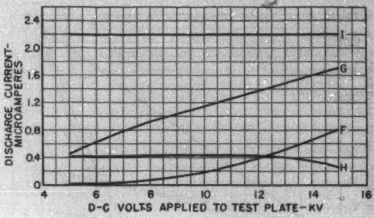


Figure 9 (left). Gap length = 1.00 inch

Figure 10 (right). Gap length = 1.50 inches



Curves H and I in the solid lines represent the discharge characteristics of radioactive preionizing neutralizers with the test plate electrified to negative polarity. Curves H and I in the dashed lines are similarly the discharge characteristics for the test plate electrified to positive polarity. Curves F and G represent the discharge characteristics of high-voltage preionizing neutralizers without proximity ground—the solid lines for the negative test plate and the dashed lines for the positive test plate

proximity ground facilities or they may function from the ionization caused by alpha emission from radioactive isotopes. In addition to the classification as shown in the illustration, the individual neutralizers are labelled with code letters for convenience of reference on the curves showing the graphical presentation of their discharge performance.

The test results portraying the characteristic discharge facilities of the neutralizing units at the indicated gap lengths are displayed in Figures 6, 7, 9, and 10. The discharge characteristics of the high-voltage preionizing neutralizers, with proximity grounds, units A, B, C, D and E, shown in Figure 6 for a 1.00-inch gap, appear to possess close linear response to the varied d-c electrification of the test plate. Actually, if the electrification of the test plate could be increased indefinitely without spark-over to the a-c neutralizers, all of the ions in the ambient air would be appropriated for discharge purposes as rapidly as they are formed and, for this reason, the rate of discharge would decrease. Hence, the curves of discharge current, as a function of the d-c electrification of the test plate, would gradually approach the horizontal as occurs for any saturation curves. This typical saturation for a-c neutralizers is shown by all of the discharge curves in Figure 7 for the 1.50-inch gap except for unit D.

The decreasing rate of discharge current with increased gap length, as disclosed by comparing the curves in Figure 7 with those of Figure 6 for a-c neutralizers, occurs as a compromise of two opposing influences. If the concentration of ions in the air near the neutralizer points remained constant, the above-mentioned decreasing rate in discharge current would vary hyperbolically as an inverse function of the gap length. Actually, when the test plate is electrified to negative polarity, the a-c ionization of the air near the tips of the ionizers becomes greater as the gap length increases. This anomalous condition occurs as the gap length increases because the negative field of the test plate opposes with less inhibiting influence the negative electric field at the points of the neutralizer on each negative half of the a-c cycle.

As disclosed by the cathode-ray oscilloscope, when the ionizing points are electrified with negative polarity during one half-cycle in comparison with positive polarity during the other half-cycle at the same maximum voltage, the ionization current through the air is increased to double or triple. Also comparative test data for an induction neutralizer at a given gap length and impressed voltage provides additional evidence of the greater ionization from negatively charged points, the test plate being successively electrified to positive and negative polarity. The discharge current for negative points in the tests on induction neutralizers ranges from two to five times higher than for points electrified under identical conditions with positive polarity. The data in the curves of Figure 8 give evidence of this induction ionization phenomenon for a single 4-mil bristle. In addition, the bristle when negative to the positive test plate, as indicated by the dashed lines, requires a lesser threshold voltage for ionization than when charged to positive polarity by some 38 per cent.

In Figures 9 and 10, the discharge current for the radioactive neutralizers, units H and I, is seen to be substantially independent not only of the range of d-c electrification employed in the tests but, also, of the lengths of air gaps utilized. The causes for this anomalous behavior may be obscure without explanation. The ions of the ionized air within the gap space, resulting from alpha bombardment, are wholly appropriated for neutralizing the impressed charges on the test plate by attraction resulting from d-c electrification as low even as five kv. Hence, the higher voltages of the d-c electrification can attract to the test plate no more ions than those provided by this constant rate of ion production, thus in this manner providing the flat discharge characteristic as shown by the solid lines. Even with the increase of gap lengths from 0.75 inch to 2 inches, the low d-c electrification of 5 kv is still found sufficient to appropriate for neutralization the ions in the gap as rapidly as they are produced.

Another interesting and, indeed, startling disclosure which pertains to ionization from alpha emitters, and a phenomenon which obeys perfectly nature's law and order, relates to the decrease of discharge current when the test plate is electrified to positive polarity. The electric field from the plate, in opposing the movement of the positively charged alpha particles, slows their speed and thus decreases their ionizing properties. For the radium neutralizer, the discharge currents for both the 0.75 and 1-inch gaps are gradually reduced from 0.25 microampere, in each instance, at 5-kv plate voltage to zero at 15 kv. The polonium neutralizer similarly shows evidence of this submission effect on the speed of the alpha particles by the positive field influence. However, this effect is displayed to a lesser degree in this instance by virtue of the greater ionization from a more intense source of alpha emission than from the radium neutralizer. These unique discharge characteristics are indicated by the curves with the dashed lines in Figure 9.

Also in Figures 9 and 10, the high-voltage a-c neutralizers without proximity grounds, see units F and G, are seen to yield strikingly low discharge currents in comparison with the discharge currents in

Figures 6 and 7 for the high-voltage a-c neutralizers with proximity grounds. Special attention is directed to the unique performance curves of unit F which provide clear evidence of a superposed component of ionization by induction to the a-c ionization for the higher values of d-c electrification.

Special tests demonstrated that grounded machine members on the opposite side of electrified stock from a neutralizer without proximity ground, contrary to general belief, provided substantially no increased effectiveness in the ionization of the neutralizer. On the other hand, however, if grounded members were located on the same side of the electrified stock as the neutralizer and within an inch or so of the ionizing points, the ionization and the discharge current caused by the neutralizer as a result of these artificial proximity grounds increased severalfold.

The discharge curves in Figures 6 and 7 show that induction neutralizers, see units K, L, M, N, and O, do not start ionizing until the threshold potential gradient of ionization is attained at voltages ranging from the order of 7 kv to 13 kv, depending on the gap length. However, when ionization once starts, the discharge current increases rapidly as an exponential function of the d-c voltage existing between the test plate and the grounded bristles of the neutralizer. The threshold voltage of ionization and the subsequent discharge current are both critically affected by the manner in which the metallic bristles are disposed along the neutralizer bar, as well as by the nearness of grounded machine members either on the same side of the electrified stock as the static bar or on the opposite side.

Grounded members near the bristles of induction neutralizers divert the electrostatic field of the electrified stock from the bristles, thus reducing the potential gradients at the tips of the bristles and, hence, decreasing the discharge current. Or, if the grounded machine members are located close to and on the reverse side of the electrified stock from the neutralizer, the increased localized capacitance between the electrified stock and ground decreases the potential of the stock and, therefore, reduces the induction and the discharge facilities of the bristles.

The electrostatic principles which are basic to the design of effective induction neutralizers seem not to have been appreciated or applied by the manufacturers of units K, L, M, N, and O. By utilizing the known principles of electrostatics to advantage, the discharge facility of these units may be increased in the order of two- to threefold, as well as improving their service by lowering the threshold voltage of ionization and thereby increasing their safety utility in hazardous applications. A manufacturer appreciating the limitations of induction neutralizers and their restricted areas of design could best serve industry by concentrating on the development of a single unit of superior performance rather than by attempting to market several units of inferior characteristics.

Proximity Effects of Ionizing Points

The spacing distance between ionizing points along a neutralizer bar is a critical element of design which influences to a marked extent the degree of ionization under established conditions and, hence, affects the magnitude of the discharge current both for high-voltage preionizing units and for induction units. If two ionizing points on a high-voltage preionizing neutralizer with proximity ground are separated by several inches, then for a constant a-c excitation of the points, constant d-c electrification of the test plate, and constant gap length, the discharge current is found to be closely double that for a single point. If, however, the two ionizing points are moved nearer and nearer together, the discharge current no longer remains double that caused by a single point but, rather, it appreciably diminishes in value. This proximity effect for two such preionizing points may be observed by reference to the solid-line curves in Figure 11.

When the preionizing points were energized to their rated a-c voltage and the test plate was electrified to a d-c voltage of 15 kv, the discharge current was 7.0 microamperes, while for two points separated by a distance of $8^{13}/_{16}$ inches, the discharge current of 13.95 microamperes shows a value of about double. After the two points were relocated to a separation of only 13/16 inch, the discharge current for the two points was reduced to 11.1 microamperes, a value which shows a reduction in effectiveness for the second point of 40.7 per cent.

In testing the proximity effect of grounded needle ionizers subjected to electrostatic induction, first a single needle was located over the test plate with the same gap length as for the ionizing points of the high-voltage preionizing unit. With the same impressed d-c voltage of 15 kv as employed in the previous proximity tests, the discharge current was 20.7 microamperes. When the two needles were used and separated by the large spacing of $8^{13}/_{16}$ inches, the discharge current of 41.1 microamperes was likewise found to be about double that for a single needle. After the two needles were readjusted to a separation of 13/16 inch, the discharge current measured 34.3 microamperes, showing a reduction in effectiveness by proximity effect for the second needle of 32.8 per cent. For a separation of one-fourth inch between the two needles at the same gap length and applied d-c voltage as used in the other proximity tests, the discharge current measured 23.0 microamperes, showing a reduced effectiveness for the second needle by proximity effect of 87.5 per cent.

Curves showing the discharge current with respect to the electrification of the test plate for the ionization by induction of the single needle and the two needles with spacings of $8^{13}/_{16}$ inches and 13/16 inch, as described above, are also included in Figure 11 for comparison, as shown by the heavy lines. In addition to disclosing graphically the specific proximity effects, as explained in the text, the two families of curves in Figure 11 provide an excellent comparison of the typical discharge characteristics for a-c preionizers with proximity ground and for induction ionizers. If the discharge curves for alpha ionizers had also been included here, when the test plate was negatively electrified, they would be characterized by practically horizontal straight lines.

This proximity effect, as the two points are brought into closer relationship one with the other, is caused by the lowered potential gradient at each ionizing point resulting from the inhibiting influence of the opposing electric field of the same polarity from the other ionizing point. The electric field of one ionizing point repels that of the neighboring point, thereby distorting the field distribution, weakening the potential gradients at the tips of the points, and reducing the ionization and the consequential discharge current.

Where a tuft composed of many metallic bristles is employed rather than a single wire bristle or a needle point, proximity effect likewise greatly reduces the discharge currents in the individual components of the tuft. In an actual test, the discharge current for a specially prepared tuft composed of 475 wire bristles, each of 4-mil diameter, for the same gap length and d-c voltage as employed in the former tests for proximity effect, was only 1.9 times higher than that for a single discharge element. On the other hand, the threshold voltage of ionization for the tuft possessed the disadvantage of being 2.34 times higher than that of a single wire bristle.

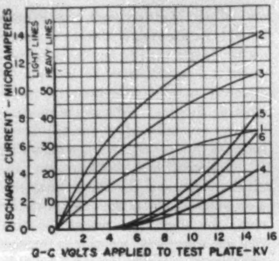


Figure 11. Curve (1) represents the discharge characteristic of a single a-c ionizing point with proximity ground, curve (2) two such ionizers $8^{13}/_{16}$ inches apart, and curve (3) the two ionizers 13/16-inch apart. Note the typical saturation effect for this type of ionizer. Curve (4) represents the discharge characteristic of a single needle induction ionizer, curve (5) two such needles $8^{13}/_{16}$ inches apart, and curve (6) the two needles 13/16-inch apart. Observe the exponential rise which characterizes the induction ionizers

Obviously the proximity effect is particularly high in any brush-type neutralizer in which the wire or other bristles are literally packed solid, such as in units L, M, N, and O; and, under this condition, in addition to passing a relatively low discharge current the threshold voltage of ionization becomes maximum.

Degree of Discharge Current Required

In the calendering of plystock for rubber tires, in the coating of textiles with rubber or plastic compositions, and in various other processing operations at low speed in which the voltage to ground of the electrified stock ranges in the order of 25 to 50 kv, the electron density lies in the general order of 10^{11} electrons per square inch of surface. The low speed at which the stock is processed in these operations ranges for the most part between ten to 50 yards per minute. Under these conditions with stock moving at, say, 20 yards per minute, the rate of movement of electrons in a foot width of the electrified stock lies in the order of 10^{13} electrons per second which, for complete neutralization, corresponds to a discharge current in the order of one microampere.

The above figure of discharge current per foot of stock width indicates that the radioactive neutralizers and the high-voltage preionizing units without proximity grounds, the performance data for which are shown by the curves of Figures 9 and 10, appear to meet more or less satisfactorily the requirements imposed on them by these industrial processes provided, of course, the units are mounted correctly and maintained properly. In other kinds of processing where voltages are encountered in the order of 100 kv or more and, especially, where the speed of the stock is several times higher than 50 yards per minute, such as in rewinding operations of nylon and of other highly insulating stock and in newspaper printing, neutralizers of these types could not be expected to yield complete neutralization of the electrified stock.

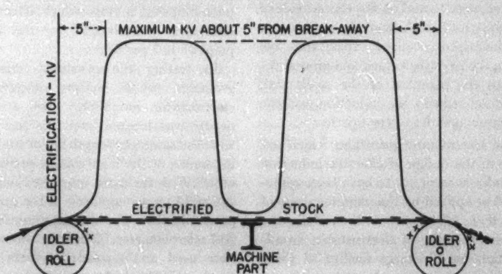
Perhaps complete neutralization is unnecessary in some instances, but the criterion which prescribes whether or not complete neutralization is necessary only becomes apparent through an analysis of the extraneous elements which are imposed by the processing operations. For example, in neutralizing electrified plystock to mitigate fire hazards, voltages below 5 kv are safe unless unusual circumstances along the route of the moving stock introduce increased capacitance values whose energy in potential discharge sparks, should they occur, lies above the threshold value required for the ignition of flammable vapor-air mixtures, such as a threshold ignition energy in the order of 0.0001 to about 0.003 watt-second for various liquid hydrocarbon products.

On the other hand, the copious a-c ionization and the high discharge currents disclosed by the data in Figures 6 and 7 indicate that the high-voltage preionizing neutralizers with proximity grounds provide ionizing facilities capable of completely neutralizing stock under even the highest industrial electrification, such as in newsprinting, the manufacture of plastic films, and the embossing of wall paper. Where such high-voltage electrification occurs, induction neutralizers of the metallic-brush type, when properly designed and constructed, also perform satisfactorily provided they are mounted correctly and maintained effectively.

A fault commonly encountered in the mounting of neutralizers for industrial applications which may render any of them ineffective is that of installing the stock at locations where a grounded machine part lies directly beneath and close to the stock. The voltage between the stock and ground is thereby lowered through the capacitor effect and, consequently, incomplete neutralization results. Hence as the stock, still electrified, progresses beyond the neutralizer, the voltage reappears at high value. This situation is illustrated in Figure 12.

The induction neutralizers, by deriving their ionizing properties as a function of the voltage between the electrified stock and ground, possess inherent facilities for providing self-regulation. The high-voltage preionizing neutralizers with proximity grounds appear from the test data to provide ionization and discharge currents far in excess of the requirements for many industrial processes. While a vast excess of ions in the ambient space around the stock after its neutralization causes no ill-effects neither does it possess any virtue. Rather than to create ionization far in excess of the need for complete neutralization, the manufacturers of the electrostatic equipment could best serve the interests of improved safety by achieving the desired result through the use of lower voltage equipment so designed as to be devoid of potential shock hazards from vicious high-voltage sparks and thus, also, to provide a maximum degree of safety from ignition of flammable fuel-air mixtures.

Figure 12. The rapid rise in voltage shown following the break-away of the electrified stock from the idler or processing rolls results from the rapid decrease of capacitor effect between the idler roll and the retreating ions on the moving charged stock. The reverse effect in the voltage is observed where the stock approaches a roll. Also the dip in voltage is caused by the localized high capacitance of grounded metal structures near the moving stock



Anomalies in Electrifying and Discharging Processed Stock

Unlike ionized air, as already considered, in which both the positive ions and the electrons, as well as the molecules, are free to move through space, the contact electrification of substances in the solid state takes place only by the mass migration of electrons from one substance to the other. From the physiology of crystalline solids, the elemental atoms are elastically constrained by electric forces to definite positions and space patterns in the latticed structures which compose the molecules of the crystalline substances. However, the atoms and the molecules continually undergo more or less vibration about their mean positions of equilibrium because of their stored thermal energy. The substances soften or melt when the increased vibrational forces resulting from applied thermal energy exceed the binding ties between the atoms originating from their inter-atomic electrical forces.

The metals, being more tightly packed with atoms and molecules, possess higher density than gases and liquids. As a result of this compact structural composition, their outermost planetary electrons possess the unique property of freely interchanging from one atom to another. These are called free electrons and the availability of them in vast numbers accounts for the relatively high electrical conductivity of the metals. In nonconductors, practically no free electrons exist, although electrons can be liberated from surface atoms by the influence of powerful extraneous electric fields, ultraviolet radiation, or X rays. Of more importance in static electrification, however, is the liberation of electrons from surface atoms by virtue of the electrical unbalance of intra-atomic forces at the interfacial boundaries between contacting areas of different kinds of substances.

When contact thus occurs, electrons migrate from the substance of higher dielectric constant to the substance of lower dielectric constant through the many points of surface contact. The electron migration continues until equilibrium of the superimposed system of electric forces at the interfacial boundaries is established. The phenomenon is correctly known as contact electrification, although frequently it is erroneously referred to as frictional electricity. Actually friction plays no part in electrification other than to present more points of contact between contacting substances, especially fibrous materials, whereby more electrons can migrate from one substance to the other and thus increase the total electric charge.

If one of the two contacting substances is a poor conductor, the electrons are impounded at the local areas on the one substance to which they migrated and when separation of the two substances takes place especially if it occurs rapidly, they remain thus impounded where they migrated. If separation occurs slowly, however, a large proportion of the electrons may wiggle back to the other substance from whence they came before final separation occurs and, hence, relatively low electrification results. Thus high-speed processing may give rise to higher electrification than low speeds on substances which are not excellent dielectrics and, mostly, the common run of stock processed in industry is not highly dielectric.

Thus, the electrification of processed stock is seen to consist either of the migration of masses of electrons into the stock, if it becomes negatively electrified, or from the stock into the processing rolls, if the stock becomes positively electrified. Since the acquired electrons or the residual positive ions in either instance reside at the surface of the stock, electrification is therefore a surface phenomenon. As already stated, processed stock is most generally electrified to negative polarity and, hence, the stock is provided with vast numbers of excess electrons, more or less uniformly distributed over the surfaces.

In the event that the stock passes between processing rolls of similar metal, like rubber plystock between the steel rolls of a calender, electrification of the same polarity is imparted to both sides of the material. Any inequality of electron density on the two sides of the processed stock becomes equalized almost instantly as the stock emerges from the rolls. This occurs under the driving influence of the high potentials of repulsion which develop between the charges of like polarity on the two sides of the stock, an equalization which not only occurs over any one surface but also takes place through the stock as a transfer of electrons from one surface to the other.

While the contact electrification of the two sides of the stock is generally of the same polarity, when the stock passes between identical steel rolls, it may differ considerably in amount on the two sides. Localized surface contaminations or variations in homogeneity of the stock on one side with respect to the other account for these differences in degree of electrification. The exception occurs in those rare instances when extraordinarily high dielectric material is processed.

As a limiting case, consider stock passing over an idler roll. The stock is obviously electrified by contact on one side only; but as tests have repeatedly demonstrated, the other side also normally becomes charged to the same polarity and intensity. That this situation occurs gives unmistakable evidence of the inherent powerful electric forces which are at work within the electrified substances to cause equalization of charge. In certain instances, such as in the printing industry, where stock in rotogravure presses passes between a rubber roll and a copper printing cylinder, the stock may be electrified by contact with opposite polarities on the two sides. Ordinarily, recombination and readjustment of the charges of opposite polarities occur through the paper, leaving both sides of about equal and like polarity. By the same token, a static neutralizer of appropriate type when located on either side of the stock discharges it regardless of whether the stock is electrified to positive or negative polarity or whether areas of opposite polarity exist on any one side.

Considerable interest and, indeed, amazing misunderstanding have centered around the use of a recently developed powder which, when sprinkled onto electrified stock, indicates by different colors those areas electrified to positive and to negative polarities. Obviously, the powder possesses no other function—and, indeed, no other is claimed for it—than demarking the polarized areas of electrification.[10] Except for certain types of research study in the science of electrostatics, the value of this powder is essentially academic.

The contact electrification of crystalline substances has been proved by painstaking research to be independent of the amount of applied pressure. However, this revealing discovery, as important as it is, possesses only scientific interest since it relates basically to normal pressures applied to optically flat surfaces.[11] Materials which are industrially processed are neither optically flat nor do they even remotely approach such a condition. For this reason, the electrification, as acquired by crystalline substances in contact, increases greatly with pressure. Such characteristic behavior is caused by the intermeshing of the microscopic roughnesses on the two interfacial surfaces which are gradually pressed into more intimate contact with increases of pressure. Through these increased areas of contact, more electrons are enabled to migrate from the substance of higher dielectric constant into the substance of lower dielectric constant.

On the other hand, while amorphous substances, of which rubber is an example, when pressed together acquire polarities strictly in accord with the relationship of their dielectric constants, they undergo increased electrification with increased pressure only up to a certain critical value. Beyond this critical pressure, the electrification thereafter remains constant regardless of continued application of pressure. Hence a curve representing the electrification of amorphous substances in contact, made by charting the electrification as a function of the normal pressure, resembles in form a typical saturation curve.

Safety Features of Neutralizers

The use of capacitive coupling between the hot central conductor of a neutralizer bar on high-voltage preionizing units and the ionizing points has already been mentioned and illustrated in Figure 1(B). This safety provision in the construction of high-voltage neutralizers is important as a means of reducing potential fatality by touching or receiving sparks from the ionizing points.[12] Maximum safety in this regard is attained by providing each point, individually, with capacitive coupling. One such neutralizer, in addition to the capacitive coupling utilizes high resistance in series with each ionizing element to control even further the inherent safety of the unit against electric shock. The alternative method which seems unwise to employ, consists of mounting all the points on a single long metal sleeve which constitutes a capacitor element extending continuously the full length of the neutralizer. The capacitance of such a long sleeve with respect to the central conductor is relatively large and sparks discharged from any one of the many points present a real hazard to the safety of operators.

Another important safety provision comprises the incorporation of high leakage reactance in transformers which energize the neutralizers. By so doing, the short-circuit current and the sparking current directly from the transformer is maintained well below potential electrocution values—the threshold alternating current for which lies in the general order of 100 milliamperes.[13]

Four transformers, one from each of four manufacturers, were tested both for short-circuit currents and for high-voltage sparking currents. These tests were conducted in each instance with the current passing through a resistance of about 1,000 ohms which value represents about minimal electrical resistance through the human body from hand to hand or hand to feet. The short-circuit currents ranged between 0.5 and 13.6 milliamperes while the currents in sparks from the transformers ranged between 3 and 70 milliamperes. The ratio of the spark current to the short-circuit current for three transformers varied from five to 20, one unit being so well designed in this respect that the spark current was as low as the short-circuit current. The spark discharges in all but one instance were characterized as high-frequency sparks resulting from the oscillatory nature of the circuits by virtue of the inductive and capacitive parameters. This phase of the subject could be investigated to advantage by the manufacturers with the objective of improving the design and construction of the transformers to restrict spark currents, should they occur, to inconsequential shock intensities.

Similarly, tests of the short-circuit current and the sparking current from the ionizing points were made when the neutralizers were being operated under their rated impressed voltages. In these tests, the resistance of about 1,000 ohms was likewise used to represent body resistance. Under short circuit, the current from an ionizing point to ground for the seven a-c neutralizers which were tested ranged from 0.04 to 0.8 milliampere and the sparking current ranged from 0.02 to 4.0 milliamperes. The ratio of the spark current to the short-circuit current for individual neutralizers in these tests varied from one-third to five.

As an additional and significant test of the sparking intensity for each neutralizer, the tip of the little finger of one hand was brought within sparking range of the ionizing points. In some instances, intense involuntary muscular response occurred. If attendants were thus shocked by sparks while working on operating machinery, they could receive serious consequential injuries by being caught in gearing, rotating or reciprocating parts, and travelling belts or stock. In other instances, no noticeable muscular reaction occurred from feeble spark discharges and, for two neutralizers, the only noticeable perception of ionization was a characteristic sensation of electric breeze near the points rather than actual sparks.

By and large, a wide spread of difference was found to exist between both the transformers and the neutralizers with regard to safety from electric shock and reflex muscular reaction. These shock hazards constitute such important life and limb considerations to operators in the hazardous locations where the units are almost always installed, adjacent to moving parts of machinery, that it seems appropriate to emphasize here the desirability of the manufacturers making a serious investigation of the electric shock characteristics of their neutralizer equipment. Perhaps, in the course of these studies, the shortcomings of certain units with regard to these phases of safety will be revealed and the necessary steps taken to correct such deficiencies.

Criteria for the Selection and Use of Neutralizers

In selecting neutralizing equipment for use in various processing operations where static electricity either presents fire or explosion hazards or introduces problems of reduced production, spoilage of product, and hazards to the safety of operators, one may be guided by certain sign posts leading to effective service and to reasonable safety. Some of the criteria by which a discriminating selection of appropriate neutralizers may be made are listed.

1. If the processing machine for which a-c neutralizers are to be selected is inherently dangerous because of rotating gears, contacting drums, reciprocating members, or other such hazardous moving parts, the neutralizers should be selected with care on the basis of specific tests to provide conclusive evidence that no electric shocks inadvertently obtained from sparks, originating at the neutralizer points, will cause involuntary muscular response that may result in physical injury to the operators.

2. Simple yet revealing tests for determining safety from electric shock may be made after the neutralizing equipment is energized. First, connect the neutralizer bar to the transformer by appropriate connectors, ground the transformer, as indicated, and also ground the proximity ground of the neutralizer, if one is provided. Second, after the transformer is connected to the proper a-c voltage, either 120 volts or 240 volts, the shock intensity of the ionizing points may then be ascertained by moving the sensitive little finger slowly toward the point. Third, if the spark which jumps to the finger causes a shock of sufficient intensity to make one jerk away his hand, such involuntary muscular response constitutes a hazard to personal safety which should preclude the use of the particular neutralizer on the processing machine.

3. If the application of neutralizers is one involving potential fire hazard, first investigate the nature, characteristics, and control of the flammable agency before attempting to apply corrective measures for mitigating the electrostatic phase of the problem. The source of the fire hazard may reside in the evolution of residual vapors from flammable solvents still contained in plasticized coatings, cements or other adhesive preparations used in the processing operations. If possible, such sources of fire hazards should be independently controlled. Since hot sparks

are generally required for ignition under such circumstances, the voltage between the stock and ground for assured safety should be reduced at least to the order of 3 kv.

4. Where potential explosion hazards exist in processing operations within enclosures, such as the drying of stock in vertical, horizontal, or festoon driers, no static eliminator should be employed from which inherent sparks occur except where the energy of such sparks is proved by test to be less than 20 per cent of the threshold ignition energy for the types of solvents utilized in the process.

5. Because of the possible insulation failure of high-voltage transformers, their cables, and the neutralizers, such inherently dangerous equipment in this respect should not be utilized in any enclosures where potential explosion hazards exist.

6. In manufacturing operations where even low electrification may result in electrostatic problems, such as the sticking together or repulsion of sheets of paper, of textile components, and the like, induction neutralizers are obviously ineffective and unsuitable for such applications and, hence, preionizing units are definitely indicated.

7. For processing operations where the speeds are generally below 50 yards per minute and where the voltage between the electrified stock and ground lies in the low order of 25 kv, practically any neutralizer on the market is sufficiently effective to provide adequate neutralization. However, where the use of induction neutralizers is desired, a residual voltage in the small amount of 500 volts may remain even in instances of neutralizers which are properly designed in accordance with scientific principles; otherwise the residual voltage may be several times higher than the above value. The small residual charge of 500 volts or less may have no significance in many operations; but in some phases of the paper and textile industries, it may result in continued difficulties or dissatisfaction of service.

8. Where radium ionizers are being considered, one should seek aid from the manufacturer as to appropriate protective measures by which to guard against possible exposure of operators either to direct or reflected radiation and, also, he should inquire from the state department of public health as to its attitude regarding the safety of the neutralizers where operating personnel are concerned. Units of this type may serve certain applications in industry, perhaps, to the exclusion of neutralizers of any other type.

9. With regard to the use of polonium neutralizers, no problem appears to exist insofar as possible injury from irradiation by beta or gamma emission is concerned, as these forms of energy are almost nonexistent. However, potential danger from the toxic properties of unprotected polonium requires specific investigation and the advice of the manufacturer regarding the proper use of protective coatings and procedures of safe handling during the mounting and servicing of the units, as well as confirmation of the safety of these units by the state department of public health. The short two-year effective life of polonium neutralizers obviously circumscribes the use of these units to a relatively small area of industrial applications. However, as stated for radium neutralizers, polonium units also may fulfill certain specialized needs of industry where the use of others may be circumscribed by space limitations.

10. The author's experience with induction neutralizers of the metallic-bristle type, when properly designed and constructed, has demonstrated that units possessing minimum possible capacitance may be safely employed in applications where solvent vapors are present, even in enclosures as for driers, without introducing fire or explosion hazards. Experience has proved that only with difficulty can sparks be drawn from the bristles and, even should sparks occur, the low capacitance of well-designed units with respect to the electrified stock attenuates the energy in such sparks to values well below the threshold of ignition.

A Suggested Recommendation for Improving Neutralizers

In the manufacture of conventional electric power machinery, electrical control apparatus, electric communications equipment, and electronic devices, the public has derived immeasurable benefit from the standardization of the myriads of products through the cooperative efforts of manufacturers' associations. In fact, the notable progress in the electrical business in no small measure is attributable to the combined efforts of manufacturers in establishing, among the many other worthy objectives, improvements in the safety of electrical equipment provided for and used in applications where hazardous conditions prevail. Clearly, the many advantages derived from manufacturers' associations in serving the best interests of industry and the public require neither philosophical discourse nor sales promotion to extol the well-known virtues and benefits of such co-operative venture.

The manufacture of static neutralizers and the supplemental equipment has not been similarly influenced by integrating forces directed toward co-operative improvements of these products. The manufacturers, individually, have developed their antistatic equipment over the years with zealous interest centered around their own methods, designs, and construction. By and large, the unique character of the equipment has attested the ingenuity of the manufacturers. Much of the equipment has been well constructed, utilizes excellent engineering design, and provides satisfactory operation. On the other hand, some of the equipment is improperly designed, possesses inherently the very fire and explosion hazards which it is provided to control, and is unsafe with regard to electric shock and potentially dangerous sparks which may cause operators to receive consequential injuries from machines to which it is applied.

It would appear that an association of the manufacturers of neutralizers, formed for the purpose of considering those phases of basic electrostatic and engineering science which relate to industrial requirements, to the nature of electrostatic problems, and to the procedures best designed to resolve the dangers and safeguard the property and personnel in hazardous areas of industrial operations, would be equally advantageous to the manufacturers and the users of neutralizing equipment.

In particular the association might well interchange helpful information on such basic concepts of industrial electrostatics as:

1. The magnitude of a-c voltages to be safely used in neutralizing equipment.
2. Various ways and means of eliminating electric shock hazards.
3. Degree of ionization necessary in various divisions of service.
4. Minimum values of discharge current variously required for effective neutralization.
5. Characteristics of the various types of stock commonly encountered in industry with regard to dielectric properties, humidity reactions, residual electrification, preconditioning, and other such physical data.
6. The influence upon discharge effectiveness of gap lengths and of other provisions in mounting neutralizers.
7. The extent and nature of grounding facilities desired on processing machines and neutralizers.
8. The importance of capacitance values in sparking phenomena.
9. Instrumentation necessary for electrostatic surveys, tests, and maintenance.
10. An investigation of various methods of appraisal in comparing the effectiveness of neutralizers.

If these and the many other related problems which are basic to industrial electrostatics could be resolved and the controlling factors of neutralizer design agreed upon, the manufacturers then could serve industry to better advantage, eliminate the mystery surrounding their individual products, and provide safety engineers and others who purchase the equipment with competent engineering data.

In this manner, the formation of just such an association by all the manufacturers of antistatic devices could do much to clarify the confusion currently existing throughout industry regarding the generation, accumulation, and control of static electricity, the modus operandi of control methods, and the degree of effectiveness and safety which can be achieved. All of this could lead to a more cooperative understanding between the manufacturers and users of antistatic equipment with regard to improved safeguards in most effectively controlling industrial electrostatic hazards.

References

1. Static Electricity in Industry, **Robin Beach.** *Electrical Engineering* (*AIEE Transactions*), volume 64, May 1945, pages 184–94.

2. Safety from Industrial Electrostatic Hazards, **Robin Beach.** *Safety* (New York, N. Y.), March 1946.

3. Dielectric Phenomena in High Voltage Engineering (book), **F. W. Peek.** McGraw-Hill Book Company, New York, N. Y., 3rd Edition, chapter 4.

4. Effects of Electric Current on Man, **C. F. Dalziel, J. B. Lagen.** *Electrical Engineering* (*AIEE Transactions*), volume 60, February 1941, pages 63–66.

5. Mechanisms of the Electric Spark (book), **L. B. Loeb, J. M. Meek.** Stanford University Press, Palo Alto, Calif., chapter 2.

6. Energy and Vision, **Selig Hecht.** *American Scientist*, volume 32, July 1944, pages 159–77.

7. Evaluation of the Beta and Gamma Radiation Due to Extended Linear Sources of Radium, **R. D. Evans.** *Journal of Industrial Hygiene and Toxicology* (Baltimore, Md.), volume 28, November 1946.

8. Report on Polonium Static Eliminator Devices, **R. D. Evans.** *Journal of Industrial Hygiene and Toxicology* (Baltimore, Md.), volume 28, November 1946, pages 243–56.

9. Elimination of Static Electricity, **Robin Beach.** United States Patent Number 2,449,972.

10. A Method for Studying Distribution and Sign of Static Charges on Solid Materials, **H. H. Hull.** *Journal of Applied Physics* (New York, N. Y.), volume 20, December 1949, pages 1,157–59.

11. Contact Electricity of Solid Dielectrics, **H. F. Richards.** *Physical Review* (New York, N. Y.), volume 22, 2nd series, 1923, pages 122–23.

12. Dangerous Electric Currents, **C. F. Dalziel.** *Electrical Engineering* (*AIEE Transactions*), volume 65, August and September 1946, pages 579–84.

13. Effect of Electric Shock on the Heart, **L. P. Ferris, B. G. King, P. W. Spence, H. B. Williams.** *Electrical Engineering* (*AIEE Transactions*), volume 55, May 1936, pages 498–515.

Discussion

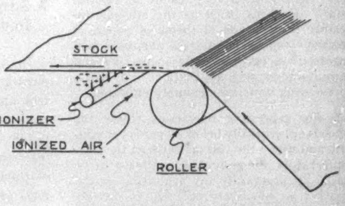


Figure 1. Preionizer without proximity ground located to provide intense ionization in the vicinity of the separation of stock from roller

Charles Carswell (Herbert Products, Inc., Woodhaven, N. Y.): Professor Beach has rendered invaluable service to the engineering profession by the preparation of this paper. His lucid explanations of the physical mechanism of static electricity and the means for its control, his listing of criteria for the selection and use of neutralizers, and his suggested recommendation for improving neutralizers, make the paper of far reaching interest and importance.

In connection with the tests made of discharge performance, it may be of value to consider further the function of the proximity ground, or other grounded conductor in proximity to the stock, in the discharging of static electricity from the processed stock.

The most important function of this conductor in the mechanism of neutralization is to provide for flow of electric current from, or to, the charged stock. The electrified stock and the grounded conductor may be likened to the two terminals of an electric switch, and the ionized air to the movable blade of such switch. When no ionization is present, the condition corresponds to the electric switch prototype with the blade open. A potential on the processed stock cannot cause current flow through the nonconductive air gap. When ionization is initiated in this gap, either from a preionizer or an induction ionizer, the switch is closed. Current flows and neutralization results.

The mechanism of current flow is the migration of masses of electrons. If the stock is negatively charged, for example, it has excess electrons on its surfaces. When ionization occurs in the air through which this stock is passing, the negatively charged electrons in the "electron avalanche" are repelled by the stock and move to the grounded conductor. The excess masses of electrons on the stock migrate from it under attraction from the positive ions, and partly combine with the positive ions and partly migrate to the grounded conductor.

Consideration of this functional relationship between the grounded conductor and the electrified stock emphasizes and explains in more detail the observation recorded by Professor Beach in connection with the tests of preionizers without proximity ground that the discharge current caused by such neutralizers increased materially when a grounded member was located near the points. It also makes clear what might otherwise appear to be an anomaly between the results from the test equipment and the practical industrial experience of neutralization with this type of high-voltage preionizers.

In the industrial applications, the high-voltage preionizers without proximity ground make use of machine parts or of supports for the preionizers, as the grounded conductor, and may prove more efficacious and more efficient than a preionizer with a grounded conductor of integral construction with it. With electrified stock passing over a metal roller, for example, the preionizer without proximity ground can be so located as to provide intense ionization in the vicinity of the separation of stock from roller as illustrated in Figure 1 of the discussion.

The roller is the grounded conductor. The neutralization of static is taking place practically at the point of origin of the charge.

Improvement of Neutralizers

Relative to the suggestion made by Professor Beach for the formation of an association of manufacturers for the purpose of considering various phases of electrostatic and engineering science, it seems that such an association would be advisable and beneficial to all concerned. However, such an association would have to be confined to the research and discussion of scientific data and its various effects on the industry as a whole. Certainly it appears very improbable, if not impossible, to expect even general standardization of static neutralizer manufacture, owing to the widely diversified fields which such manufacturing serves. In many instances, each application represents its own engineering problems, with the basic principles the same of course, but with complete and satisfactory solution hinging on consideration of particular elements peculiar only to the specific case. Conforming to rigid standards of manufacture and design would very likely hamper expeditious solutions in such instances.

However, an association of manufacturers would certainly be of unquestionable value to its member firms for the purpose of collective investigation of the many suggestions on which there is still no clear general conception. For example, those suggested by Professor Beach. In addition to the items enumerated for such research, the thermal effect on static elimination would be an item for investigation to determine the extent and benefit of various heat units in promoting neutralization of static.

E. J. Rutan (American Cyanamid Company, New York, N. Y.): Even though the electrostatic method of generating electricity is the simplest, it is the one we know least about quantitatively. Those who have encountered this problem in practice realize the lack of definite information for protecting installation so that they can be definitely considered safe.

Professor Beach has described a method for measuring the performance of electrostatic protective devices. The fact that he has developed a method for measurement is, in my opinion, the reason this paper is so important. Here is a means for grading the different devices being manufactured and used. This information is certainly most important to the so-called consumer.

In the section on "A Suggested Recommendation for Improving Neutralizers" Professor Beach makes several recommendations. On one of these I feel constructive action could be taken:

> "The manufacturers, individually, have developed their antistatic equipment over the years with zealous interests centered around their own methods, designs, and construction. By and large, the unique character of the equipment has attested the ingenuity of the manufacturers. Much of the equipment has been well constructed, utilizes excellent engineering design, and provides satisfactory operation. On the other hand, some of the equipment is improperly designed, possesses inherently the very fire and explosion hazards which it is provided to control, and is unsafe with regard to electric shock and potentially dangerous sparks, which may cause operators to receive consequential injuries from machines to which it is applied."

Here is a situation which certainly requires action by the AIEE through a technical committee. Professor Beach has suggested action along these lines. It seems to me with a method of measurement of performance available as described in this paper, it is certainly time for the AIEE to set up a committee which will include both manufacturers and potential consumers to at least set standards which will label the unsafe devices.

Induction Preheating of Electrolytic Tin Plate for Flow Brightening with High-Frequency Rotating Equipment

WILLIAM T. THOMAS
ASSOCIATE MEMBER AIEE

MANY producers of tin plate, due to its large demand, are considering new or additional equipment. In order to flow brighten tin plate at high speeds, it is necessary to use high-frequency induction heating equipment. The other methods for producing tin plate, using combustible gases as a source of heat supply or resistance heating, are at present impractical at high speeds. To the author's knowledge to date, a maximum line speed of approximately 600 feet per minute has been attained by resistance heating. Current is the limitation for obtaining high speeds by this method. By gas heating the maximum speed that has been attained is below that of resistance heating. Space is the speed limitation for gas heating.

By induction heating, speeds up to 1,400 feet per minute have been attained and this limit is now being extended to 2,000 feet per minute. Until recently the high-frequency power was supplied by oscillator equipment. To increase production, several manufacturers of tin plate have purchased and are installing preheating equipment to the oscillator equipment, consequently increasing the line speed. Oscillator equipment itself can supply the preheating power, but 9,600-cycle high-frequency rotating equipment is given preference as the preheating induction heating power supply for several reasons:

1. First cost is about 45 per cent less.
2. The over-all efficiency is better.
3. Cost for maintenance is much less. No tube replacements are required. No special highly trained personnel is needed for conventional rotating machines and their associated equipment.
4. The insulation problem is not nearly as great because of low voltage power supply.
5. Oscillator tube life can be increased by reducing peak operating oscillator plate voltage and plate current and allowing the high-frequency generator to furnish some of the needed power, if maximum line speed has been reached.

Note that the last reason can apply to any kind of additional preheating equipment other than electronic equipment.

Equipment

A simplified diagram, Figure 1, shows the location of the preheat section in the flow brightening line.

A typical construction of the preheat induction section is shown in Figure 2. This coil assembly is rated 300 kw, 440 volt, 9,600 cycles, single phase. The power is supplied by two 150-kw water-cooled motor-generator sets, one of which is shown in Figure 3. For each high-frequency machine there is an amplidyne generator control panel as shown in Figure 4. Each 150-kw induction section, see Figure 5, is composed of 12 coils connected in multiple. The last four coils in each section can be switched by contactors. Capacitors are used for power correction and are connected across the coils with flexible stranded cables. The capacitors are distributed so as to keep the circulating current at a minimum. The permanent capacitors are rated 300 kilovars each, whereas the switched capacitors are each rated 100 kilovars.

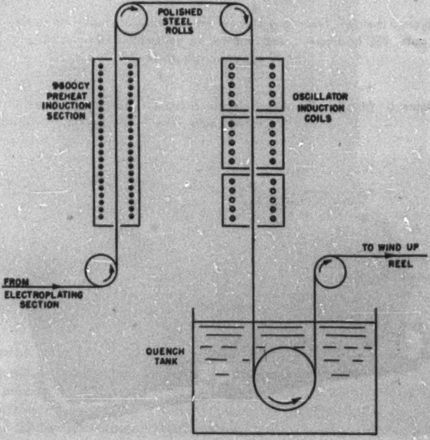


Figure 1. Simplified diagram for induction heating of continuous moving strip

---

Paper 52-4, recommended by the AIEE Electric Heating Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted August 24, 1949; made available for printing November 13, 1951.

WILLIAM T. THOMAS is with the General Electric Company, Schenectady, N. Y.

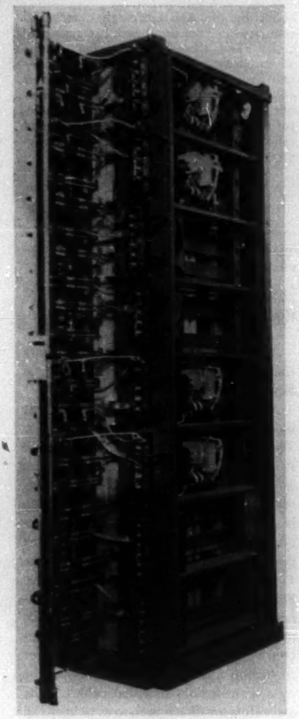


Figure 2. Coil assembly for induction heating of moving strip. View showing header tubes, capacitors, and coil switching contactors

Coil switching contactors are located in the circuit so that they are required to supply only power current. The power is supplied to the coils by water-cooled copper tubing headers.

The coils are assembled in a wooden rack, see Figure 6, in groups of three for ease of assembly and replacement and slide into the large wooden rack forming the assembly as shown in Figure 2. The turns are held in place by circular grooves in the wooden supports. The inside of the coil is protected from the strip by high-temperature resistant textolite, or micalex.

The textolite is held in place by screwed clamps. Adjacent coils are wound opposite hand to each other so that the maximum voltage between coils is only turn voltage. Each coil is connected to the headers by copper tube jumpers so that the three coil unit assemblies can be removed without disrupting the installation.

The induction coils and capacitors are connected in multiple hydraulically as well as electrically. Figure 7 shows a typical water-cooling system. Heat exchangers are used to keep the water costs at a minimum and also supply a closed, clean water-cooling system to the generators, heating coils and capacitors. Due to the water temperature requirements of the capacitors, a refrigeration system is sometimes needed for summer operation in certain parts of the country.

Figure 4. Generator control panel for induction heating control

Theoretical and Experimental Results

The first full scale tests were conducted at a large steel mill company. The development coil used is shown in Figure 8. The results from these tests are shown in Figures 9 and Table I. These results checked very well with the theory previously developed.[1] The following equations were used for calculating purposes:

Figure 3 (below). Frequency-changer set consisting of: induction motor, 2 pole, 250 horsepower, and inductor a-c generator, 320 pole, 150 kw, 9,600 cycles, single phase

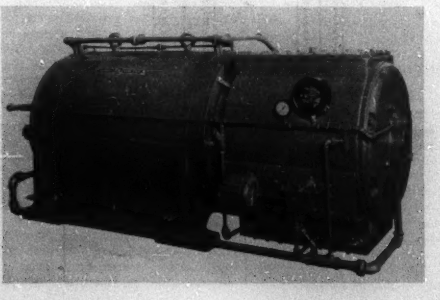
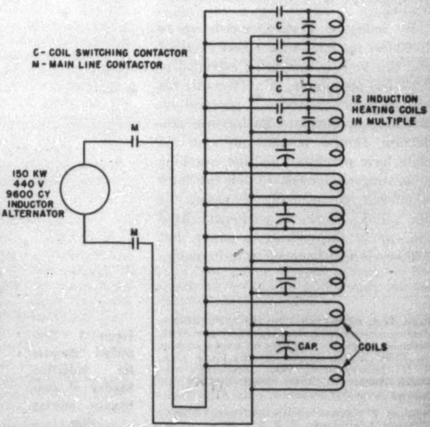


Figure 5 (right). Simplified electric circuit for induction preheat of moving strip



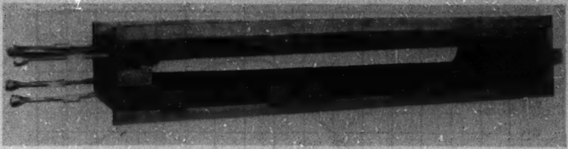


Figure 6. Coil for induction heating of moving strip

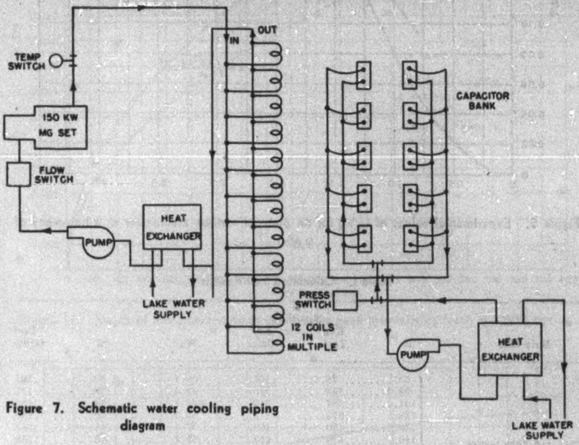
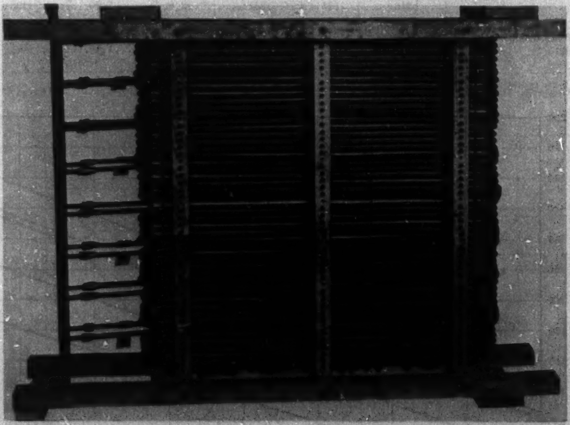


Figure 7. Schematic water cooling piping diagram

Figure 8. Developmental 9,600-cycle inductor coil used for flow-brightening of electrolytic tin plate



$$W_w = 1.62\times 10^{-8} fHt1_c b_s G(kt) \text{ watts} \quad (1)$$

$$W_e = 2.20\times 10^{-7} 1_c b_c H^2 \sqrt{f} \text{ watts} \quad (2)$$

$$(VI)_a = 2.5 K_L f H^2 h_c 1_c b_c \times 10^{-8} \text{ volt amperes} \quad (3)$$

$$(VI)_t = 5HfB\frac{t}{2} b_s 1_c \times 10^{-8} \text{ volt amperes} \quad (4)$$

In equation 4 substitute δ_f for $t/2$ when $t/2$ is greater than δ_f.

Refer to the list of Nomenclature and to Appendix I for derivation of equation 1 and the function of $G(kt)$.

The first preheating equipment was installed in a large steel mill company not far from Pittsburgh. Figure 10 shows the actual test results obtained for each 150-kw generator. A close approximation for determining the tin reflow line speed is to use 1 foot per minute per kilowatt of induction heating equipment. Thus the line speed can be increased about 150 feet per minute for each 150 kw of preheat power.

Interesting to note is that for the magnetizing force range and frequency that is used, the plot of power versus weight per base box turns out to be a straight line.

For persons not familiar with weight per base box, a rule of thumb for approximate thickness of strip is to add eight to the weight per base box and multiply by 10^{-4}. For example, 85 pound tin plate is 0.0093 inch thick.

The dotted curve for 11 coils in the circuit is calculated from equations 1, 2, 3, and 4 and compares very well with the test results. The purpose of switching coils in and out of the circuit is to fully load each generator. The number of coils required is dependent upon the strip thickness to be heated. The thinner the strip, the greater the number of coils required. All of the flux is not utilized within the sheet. Figure 11 shows the current depth of penetration at 9,600 cycles in the steel strip plotted against peak magnetizing force. In the tests conducted 150 kw, 9,600 cycle of preheating power was used with 1,200 kw, 200 kc of electronic power. The tin plated steel strip enters the induction coils at a temperature of approximately 125 degrees Fahrenheit and is heated to a final temperature of approximately 485 degrees Fahrenheit. The tin on the strip actually flows at 453 degrees Fahrenheit but usually a 10 per cent allowance is made to insure complete reflow. Therefore the strip temperature rise of 40 degrees Fahrenheit (125 to 165 degrees Fahrenheit) in the preheat section is a small portion of the total temperature rise for reflowing tin. This fact is advantageous for the use of 9,600 cycles as preheat power. The mean electrical resistivity of steel for preheat calculations is 14 by 10^{-6} ohm-centimeter compared to a mean resistivity of 22 by 10^{-6} ohm-centimeter if the steel were heated to reflow temperature.

This means that at the higher temperature the depth of penetration is 25 per cent greater.

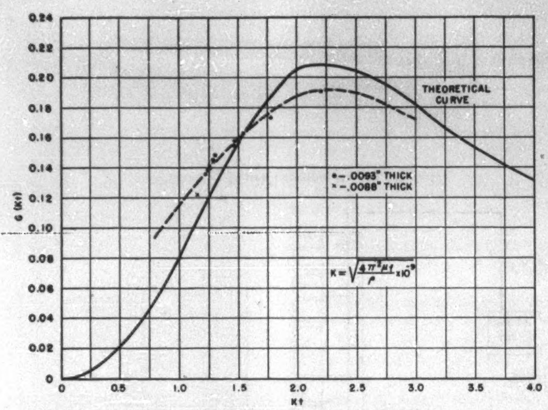


Figure 9. Experimental values of G(kt) for tin plate of various thicknesses at a frequency of 9,600 cycles

Table I. Calculated Test Results

Strip Size	H	$K_L = 0.95$ Kva Test	Kva Calc.	W_W*	kt	Test G(kt)
$27^{9}/_{16}\times0.0093$ inches	63	76.2	101	29.7	2.20	.191
	98.4	193	210	43.4	1.78	.173
	131	307	342	53.7	1.54	.165
	138	373	375	53.7	1.50	.155
	149	430	428	57.1	1.45	.157
	146	422	413	57.1	1.46	.158
	145.5	421	411	57.1	1.46	.158
	186	654	636	69.0	1.29	.149
	190	675	661	69.0	1.28	.146
	206	783	765	71.0	1.23	.139
	211	821	800	71.0	1.22	.136
	216	855	834	73.8	1.20	.138
$29^{1}/_{4}\times0.0088$ inches	172	575	555	63.0	1.28	.145
	179	616	596	63.0	1.25	.140
	213	843	850	65.0	1.15	.121
	213	843	850			

* 2-kw radiation and convection loss.

Efficiency

In order to efficiently heat the strip, it is necessary to operate at relatively low magnetic field intensities. For normal operation, a field strength of 180 to 190 oersteds is used. The depth of penetration at these values is about 0.019 centimeter or 0.0075 inch. Therefore, the thinner the strip, the more the flux is cancelled. The largest quantity of tin plate produced is from 0.009 inch to 0.010 inch thick. Even though there is some flux cancellation, tin plate can be heated efficiently at 9,600 cycles. A power density of about 30 kw per foot of coil length is supplied to 30 by 0.010 inch thick strip. Figure 12 shows the high-frequency efficiency (all losses beyond generator) for heating tin plate. The losses include I^2R of coil, capacitor losses, convection, and radiation losses. From heat transfer data literature available[2] the radiation and convection losses in the preheat section are about 0.5 watt per square inch. For over-all efficiency (actual 60-cycle power to Btu's in strip), multiply the values of Figure 12 by 78 per cent which is the efficiency of the frequency changer set supplying the high-frequency power.

Current Distribution in Strip

To give a clearer picture of what happens to the flux within the sheet, Figure 13 is plotted showing current density versus strip thickness. Referring to equation 24 in Appendix II, Figure 13 is plotted from the terms within the bracket.[3] The relative current density at the surface of a very thick strip is taken as unity. The current depth of penetration within

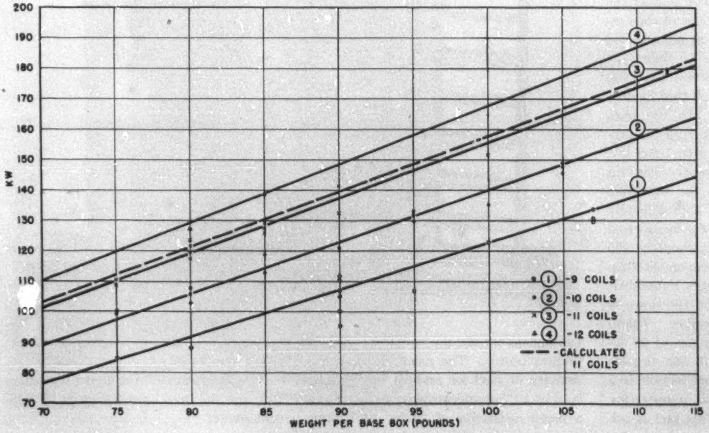


Figure 10. Tin reflow at 9,600 cycles based on 30-inch wide strip

the strip is defined as the value where the current density has dropped to $1/e$ or approximately 36 per cent of its value at the surface. The relative power density varies as the square of the current density so that at a depth equal to the depth of penetration of the power density is 13 per cent of the density at the surface. The current density lags in phase as it goes deeper into the material and it is possible for it to be 180 degrees out of phase or in phase with the surface current on the other side of the strip. This fact explains a value of greater than unity at the surface of the strip for $t/d=2$.

Limitations and Conclusions

As the strip gets thinner, the induced currents on each side of the strip tend to cancel each other and not as much power can be coupled into the strip. Consequently, it is necessary to switch coils in or out of the circuit, depending on the weight per base box, to keep the high frequency generator fully loaded. Therefore a long heating coil is required. This is the only disadvantage that high frequency rotating equipment has compared to radio frequency power. There is no reason why 9,600 cycles cannot be used as complete reflow power for speeds up to approximately 1,000 feet per minute although the high frequency efficiency will be reduced several points below that of Figure 12.

There is no doubt from the facts presented that 9,600 cycles can be used as power for heating electrolytic tin plate most effectively. Its superiority has been proven in efficiency of operation and in economy both in initial cost and in maintenance.

Nomenclature

W_w = total watts induced in strip
f = frequency, cycles/second
H = peak magnetizing force, oersted, $0.4\pi nI\sqrt{2}$
t = strip thickness, centimeters
l_c = length of coil, centimeters
b_s = width of strip
k = a parameter
$G(kt)$ = a function of (kt)
W_c = copper loss in coil, watts
$(VI)_a$ = coil volt-ampere due to flux in air
$(VI)_i$ = coil volt-amperes due to flux in iron
b_c = width of coil, centimeters
h_C = smaller dimension of coil opening, centimeters
K_L = voltage factor of coil. Approaches unity for a long coil
B = flux density (gauss); $B = 18{,}000$ gauss
μ = effective magnetic permeability of strip
I = coil current, rms amperes
n = turns per centimeter in coil
δ_s = depth of flux penetration, centimeters

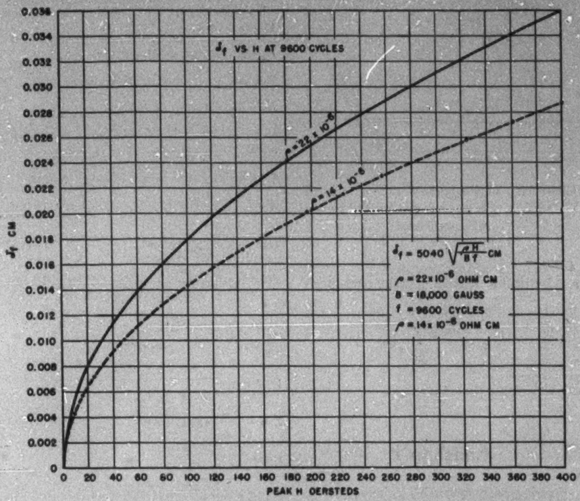


Figure 11. Depth of flux penetration in steel versus peak magnetizing force at 9,600 cycles

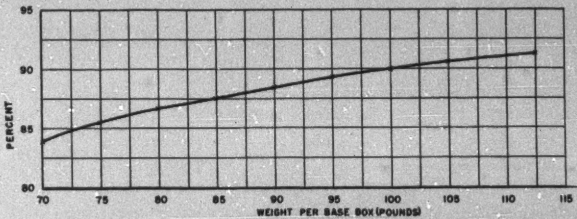


Figure 12. High-frequency efficiency of preheat section (includes all losses beyond generator)

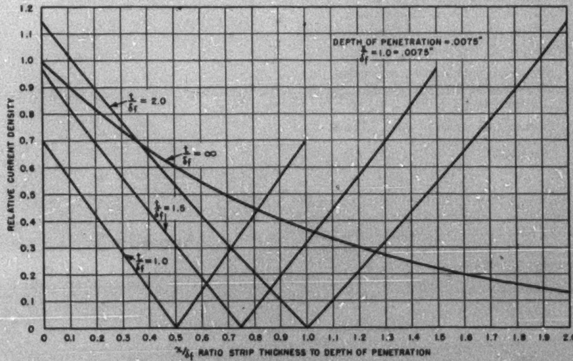


Figure 13. Curve showing current distribution through steel strip for various thicknesses of the strip

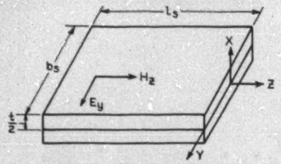


Figure 14. Showing the direction of induced H and E in strip using a right-handed system of co-ordinates

ρ = electrical resistivity of strip, ohm-centimeters
E = electric field strength
D = displacement intensity
J = current density
σ = conductivity
p = charge density
ω = radians per second $\omega=2\pi f$

Appendix I

Starting with Maxwell's field equations in the differential form and using the rationalized meter-kilogram-second (mks) system except where otherwise specified:

$$\nabla\cdot\boldsymbol{D}=p \tag{5}$$

$$\nabla\cdot\boldsymbol{B}=0 \tag{6}$$

$$\nabla\times\boldsymbol{E}=-\frac{\partial\boldsymbol{B}}{\partial t} \tag{7}$$

$$\nabla\times\boldsymbol{H}=\boldsymbol{J}_C+\boldsymbol{J}_D \tag{8}$$

$$\boldsymbol{J}_C=\sigma\boldsymbol{E} \tag{9}$$

Making the following assumptions:

1. Neglect displacement current J_D.
2. Assume steady state a-c conditions with sinusoidal variations.
3. Assume constant permeability.
4. Assume an Z component of H and a variation of H in the X direction only.
5. Assume the coil produces a current sheet of I ampere per-unit length.
6. The coil is sufficiently long so that the flux density is the same as if the inductor were infinitely long.
7. Permeability[1] may be taken as equal to

$$\mu=\frac{1.8\times 18{,}000}{H}$$

Figure 14 shows the direction of induced H and E in strip using a right-handed system of co-ordinates, and Figure 15 the relative position of induction heating coil current sheet to strip.

With the assumptions equations 5, 6, 7, 8, and 9 reduce to:

$$\frac{\partial E_y}{\partial x}=-j\omega\mu H_z \tag{10}$$

$$\frac{\partial^2 H_z}{\partial x^2}=j\omega\mu\sigma H_z=\alpha^2 H_z \tag{11}$$

where $\alpha^2=j\omega\mu\sigma$. The solution of equation 11 is

$$H_z=A\cosh\alpha x+M\sinh\alpha x$$

But due to symmetry $M=0$, therefore

$$H_z=A\cosh\alpha x$$

at $x=t/2$, $H_z=A\cosh\alpha t/2=H_a$, a constant in air. Also at

$x=t/2$, $H_a=I_{\text{peak}}$ amperes per-unit length

$$A=I_{\text{peak}}/\cosh\alpha t/2$$

$$H_z=\frac{I_{\text{peak}}}{\cosh\alpha t/2}\cosh\alpha x \tag{12}$$

Substituting equation 12 in equation 10 and integrating equation 10, evaluating at $x=t/2$.

$$E_y=-\frac{j\omega\mu I_{\text{peak}}}{\alpha}\tanh\alpha t/2 \tag{13}$$

let $k=\alpha$.

$$Z=-\frac{E_y}{I_{\text{peak}}}=\omega\mu t G'(kt) \tag{14}$$

where

$$G'(kt)=\frac{1+j}{2kt}\left[\frac{\sinh kt+j\sin kt}{\cosh kt+\cos kt}\right]$$

$$\text{Power}=\frac{1}{2}I_{\text{peak}}^2 Z_{\text{real}} \tag{15}$$

Converting from mks system to centimeter-gram-second system and using the real portion $G\,(kt)$ of $G'(kt)$ the power equation 15 reduces to:

$$\text{Watts/cc}=\frac{1}{2}\mu f H^2\frac{1}{2kt}\times\left[\frac{\sinh kt-\sin kt}{\cosh kt+\cos kt}\right]\times 10^{-7} \tag{16}$$

where

$$k_{\text{real}}=\frac{1}{\delta_f}=\sqrt{\frac{\omega\sigma\mu}{2}}=\sqrt{\frac{4\pi^2\mu f\times 10^{-9}}{\rho}}$$

Let l_c = length of coil (cm)
b_s = width of strip (cm)
t = thickness of sheet (cm)

Then

$$W_w=\frac{1}{2}\mu H^2\cdot f\cdot(t\cdot l_c b_s)\cdot G(kt)\times 10^{-7}\text{ watts} \tag{17}$$

Letting $\mu=1.8\times 1{,}800/H$.

$$W_w=1.62\times 10^{-8} f\cdot H\cdot t\cdot l_c\cdot b_s\cdot G(kt)\text{ watts} \tag{18}$$

Appendix II

Starting with equation 11 in Appendix I

$$\frac{\partial^2 H_z}{\partial x^2}=j\omega\mu\sigma H_z=\alpha^2 H_z \tag{19}$$

solving this equation in terms of exponential function

$$H_z=A\epsilon^{-kx}\cdot\epsilon^{-jkx}+M\epsilon^{kx}\cdot\epsilon^{jkx} \tag{20}$$

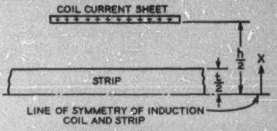


Figure 15. The relative position of induction heating coil current sheet to strip

Using the same assumptions as in Appendix I it can be shown that:[2]

$$J_y=\sigma E_y=C(1+j)[\epsilon^{-kx}\cdot\epsilon^{-jkx}] \tag{21}$$

where C is a constant

$$J_y\text{ is proportional to }\epsilon^{-kx} \tag{22}$$

where

$$k=\frac{1}{\delta_f}\quad\text{and}\quad\delta_f=5{,}040\sqrt{\frac{\rho}{\mu f}}\text{ centimeters}$$

Equation 21 applies if the strip is very thick. For thin strip, the constant M is not zero in equation 20. If the strip were excited from one side, equation 21 becomes:

$$\begin{aligned}J_y=C(1+j)[&\epsilon^{-kx}\cos kx+\\&\epsilon^{-k(2t-x)}\cos k(2t-x)-\\&j(\epsilon^{-kx}\sin kx+\epsilon^{-k(2t-x)}\times\\&\sin k(2t-x)]\div(1-\epsilon^{-2kt}\cos 2kt+\\&j\epsilon^{-2kt}\sin 2kt)\end{aligned} \tag{23}$$

However, the thin strip is excited from both sides in an induction heating coil. These induced currents oppose each other and from symmetry the current density at the exact center of strip will be zero. By super position, substituting $t-x$ for x in equation 23 and subtracting these two intensities, the total electric intensity becomes:

$$\begin{aligned}J_y=J_{y1}-J_{y2}=C(1+j)\{&\epsilon^{-kx}\cos kx+\\&\epsilon^{-k(2t-x)}\cos k(2t-x)-\\&\epsilon^{-k(t-x)}\cos k(t-x)-\epsilon^{-k(t+x)}\times\\&\cos k(t+x)+j[\epsilon^{-kx}\sin kx+\\&\epsilon^{-k(2t-x)}\sin k(2t-x)-\epsilon^{-k(t-x)}\times\\&\sin k(t-x)-\epsilon^{-k(t+x)}\sin k(t+x)]\}\div\\&(1-\epsilon^{-2kt}\cos 2kt+j\epsilon^{-2kt}\sin 2kt)\end{aligned} \tag{24}$$

As the strip thickness increases, equation 24 approaches equation 1.

References

1. INDUCTION HEATING OF MOVING MAGNETIC STRIP, R. M. Baker. *Electrical Engineering (AIEE Transactions)*, volume 64, April 1945, pages 184–89.

2. THE BASIC LAWS AND DATA OF HEAT TRANSMISSION, W. J. King. *Mechanical Engineering* (New York, N. Y.), March-August, 1932.

3. THEORY AND APPLICATION OF RADIO FREQUENCY HEATING (book), C. H. Brown, C. N. Hoyler, R. A. Bierwirth. D. Van Nostrand Company, Inc., New York, N. Y., 1947, pages 45–62.

No Discussion

An Electronic Apparatus for the Study of the Human Operator in a 1-Dimensional Closed-Loop Continuous Pursuit Task

C. E. WARREN
MEMBER AIEE

P. M. FITTS
NONMEMBER AIEE

J. R. CLARK
ASSOCIATE MEMBER AIEE

THE quantitative understanding of the behavior of a human being functioning as an element of a closed-loop control system has become a matter of considerable importance in recent years. In addition to the purely military problems of high-speed flight, fire control, and similar tasks there are many situations such as automobile driving, operating machine tools, and other complicated mechanical devices, where a specific knowledge of human characteristics would make possible improvements in design and performance of the system. The study of human behavior in this area has been carried on by psychologists and industrial engineers for some time. However, the rapid development of the theory and practice of feedback control systems in the past several years has given new impetus and interest to the study of the nature of human behavior in analogous situations.[1–6] See reference 8 for a very complete bibliography on subjects of interest in this field.

It is evident that the problem of establishing a theory of human behavior in feedback systems is a large and complicated matter. While a considerable amount of research has already been done in this area it is not certain at this time that the fundamental questions relative to the problem have been posed. A few of the questions that are under consideration now are listed as follows:

1. Is the human a linear system under certain conditions and if so what are the specifications of these conditions?
2. How does human reaction time enter into a seemingly continuous task?
3. What part of the human response is random and what part may be attributed to a systematic mechanism?
4. Is the human essentially a continuous or an intermittent system, or perhaps at times continuous and at times intermittent?

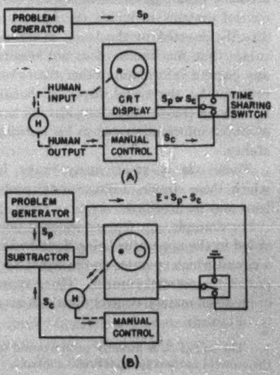
Many other questions may be asked, each in turn raising new problems. It becomes apparent that considerable fundamental research with the objective of determining the nature of the problems relative to the behavior of humans in closed-loop control situations is required. This paper describes an electronic apparatus which produces a simple, 1-dimensional, visual input, manual output control situation. This equipment is being used in exploratory studies of the behavior of humans as an element in a feedback system.

A simplified block diagram of the apparatus is shown in Figure 1. The input to the human is visual and is provided by a cathode ray tube (CRT) display. The operator sees a dot (target) and a circle (cursor) on the CRT display, his task being to keep the circle superimposed on the dot. The output of the human is a movement of the fingers, wrist, or arm or more generally any possible manual, adjustive characteristics of the human.

Two general types of 1-dimensional problems are provided. The first, which is designated as following pursuit, may be described with reference to Figure 1(A). A signal voltage S_p of prescribed time variation is generated by the problem generator. This signal is fed into the CRT display through the time sharing switch where it causes a corresponding motion of the dot on the face of the cathode ray tube. A second signal voltage S_c is created by the humans' manipulation of the manual control and is fed into the display through the time sharing switch where it produces a corresponding motion of the circle. The basic task of the human operator is to keep the circle superimposed on the dot.

For the second type of problem, designated as compensatory pursuit, the essential features are shown in Figure 1

Figure 1. Block diagram showing circuit connections for following and compensatory types of pursuit problems



Paper 52-8, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted May 21, 1951; made available for printing November 13, 1951.

C. E. Warren, P. M. Fitts, and J. R. Clark are all with Ohio State University, Columbus, Ohio.

The work described in this paper was supported in part by the United States Air Force under contract AF 33(038)-10528, monitored by the Perceptual and Motor Skills Research Laboratory, of the Air Training Command Human Resources Research Center with the Ohio State University Research Foundation, Columbus, Ohio. Permission is granted for reproduction, translation, publication, use and distribution in whole or in part by or for the United States Government.

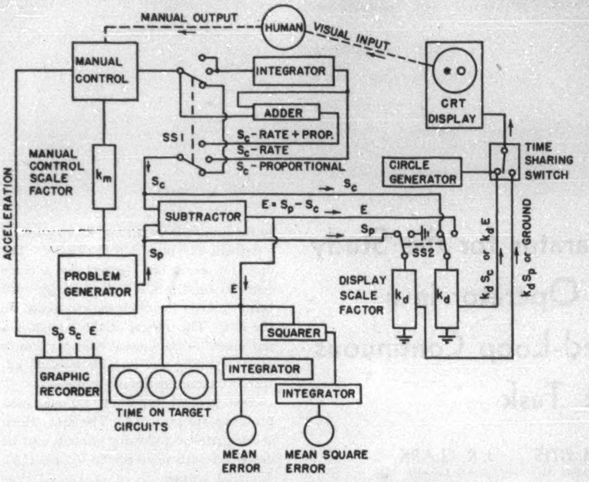


Figure 2. Block diagram of electronic pursuit apparatus

(B). Here the voltage signals S_p and S_c are created as before but they are first fed into a subtractor where the difference $S_p - S_c = E$ (error) is formed. The error signal E is fed to the time sharing switch which is now connected so that the dot is stationary at the center of the display tube and the error signal causes a proportional motion of the circle. Again the basic problem is to keep the circle superimposed on the dot.

Description of Apparatus

The following and compensatory pursuit types of problems described above are each examples of simple, closed-loop control systems in which the human is directly inserted in the loop. This assumes that the human does not receive any current information other than that provided by the visual display although he may, of course, make considerable use of stored information due to his learning ability.

There are a great many ways in which these simple, 1-dimensional problems may be modified.

For example, one feature which is provided in the apparatus being described is a circuit which operates on the signal output of the manual control. Three types of modified manual control signals S_c may be obtained.

1. The signal S_c is directly proportional to the control motion (proportional control).
2. The signal S_c is proportional to the time integral of the control motion (rate control). In other words if the control is displaced a fixed amount the signal S_c will increase linearly with time with a slope proportional to the amount of control displacement.
3. The signal S_c is a linear combination of proportional and rate control (aided-rate control).

Several important general features were taken into consideration in the design of the electronic pursuit apparatus. First of all it was decided that the apparatus should provide automatically several scores which could be used as a direct measure of performance on each test run. This is necessary if large quantities of data for statistical analysis are to be accumulated. Secondly, the apparatus should provide quickly and with a minimum of adjustment a variety of different types of problem conditions to the subject. Also it was decided that the apparatus should be constructed basically so that various kinds of modifying elements could easily be inserted into the feedback path.

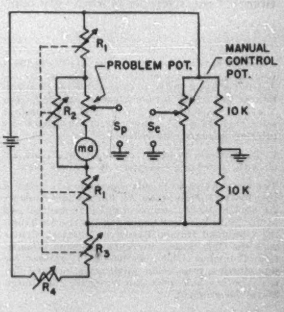
Now that a general description of the nature of the electronic pursuit apparatus has been given, a more detailed explanation can be approached with the aid of the block diagram of Figure 2. Space limitations prevent the presentation of all details of the circuitry.

In the upper left-hand corner of Figure 2 is shown the manual control. Selector switch *SS1* determines whether the control signal is to be of proportional, rate, or aided-rate type. From *SS1* the control signal S_c enters the subtractor where it is subtracted from the problem generator signal S_p to form the error signal E. All three signals S_c, S_p, and E are carried to selector switch *SS2* where either the compensatory or following type of pursuit problem is selected.

With switch *SS2* in the position shown a following pursuit problem is set up. In this case the signals S_p and S_c are fed into the display scale factor circuits where the signals are multiplied by the scale factor k_d (k_d is less than one). From this point the signals enter the time sharing switch which feeds them alternately to the horizontal deflection amplifier of the oscilloscope. In the case of compensatory pursuit *SS2* is moved to the right-hand contacts so that the error signal E is fed into one of the scale factor circuits, the other scale factor circuit being grounded at its input. With this arrangement the dot is stationary on the CRT while the circle is caused to move in proportion to the signal $k_d E$.

The circle on the CRT display is created by feeding a small 20-kc voltage into the signal path as shown on the diagram (circle generator). The vertical deflection amplifier is connected to the horizontal deflection amplifier through a high-pass filter so that when the time sharing switch is in its left-hand position a small 20-kc signal is presented to both the horizontal and vertical deflection plates of the oscilloscope. By adjusting the phase shift of the filter, the magnitude of the 20-kc signal and the oscilloscope gain, a circular pattern is produced. When the time sharing switch is in its right-hand position only a dot results, since the 20-kc signal is not coupled to this channel.

Figure 3. Diagram of manual control scale-factor circuit



In the lower, left-hand portion of Figure 2 are shown the several measuring functions which have been included in the electronic pursuit apparatus. One set of circuits produces a measure of the mean square error of the operator's performance on a particular pursuit task. This meter consists of a circuit which computes the time integral of the squared error. For a known time-duration of the problem the meter may be calibrated in rms, angular error of the manual control motion. In addition to this measure the mean error which is the time integral of E, is also measured automatically.

The time-on-target circuits measure the total time that the absolute value of the error signal is less than given preset amounts. Three separate clocks are provided so that three different error bandwidths may be measured. The time-on-target circuits and the squaring circuit will be described in more detail in the following section.

The manual control scale factor which is shown schematically as an interconnection between the manual control and the problem generator provides a simple means of adjusting the amount of angular motion of the manual control required to give a full output signal. In other words, the manual control may be set to give 100 per cent output signal for 10 degrees of motion, 45 degrees of motion, et cetera. There are a number of possible circuit arrangements which will provide an adjustment of this type. The particular circuit which is used in the electronic pursuit apparatus is shown in Figure 3. The problem and manual control signals are derived from 10,000 ohm potentiometers. The resistanced R_1, R_2, and R_3 are adjusted by means of a 4-pole selector switch so that the current through the problem potentiometer is held constant at one milliampere while the current through the manual control potentiometer takes on values proportional to the scale factors desired. Resistance R_4 is used for fine adjustment of the milliammeter current to one milliampere. With this arrangement single-sided, signal voltages S_p and S_c proportional to the motion of their respective potentiometers are derived. The steps in resistors R_1, R_2, and R_3 may be predetermined to give any number and size of ratios desired.

The several major components shown on the block diagram of Figure 2, such as the integrators, adder, subtractor, and display scale factor are more or less standard electronic circuits.[9,10] The integrators are 3-stage, direct-coupled, integrating amplifiers with negative feedback stabilization. Since the mean error and mean square error meters are reset to zero before each run only moderate long time stability is required. The time constants and signal levels of the integrators are such that runs of 2 or 3 minutes may be made with high accuracy. The adders are the electronic analogue of parallel resistance adders. Subtraction is achieved by first obtaining the negative of S_c and then adding to the problem signal S_p. The display scale factor circuits are simple cathode followers driving resistance dividers which determine the value of the scale factor k_d.

All of the electronic circuitry has been worked out with two power supply voltages, a regulated (B+) voltage of 210 volts and tube regulated (B−) voltage of 105 volts. The 100 per cent signal level was chosen to be 5 volts, plus or minus, and all circuits are adjusted to have a linear, dynamic range of at least plus or minus 20 volts. After a one hour warm-up the long-time drift at critical points in the circuit is of the order of ±50 millivolts.

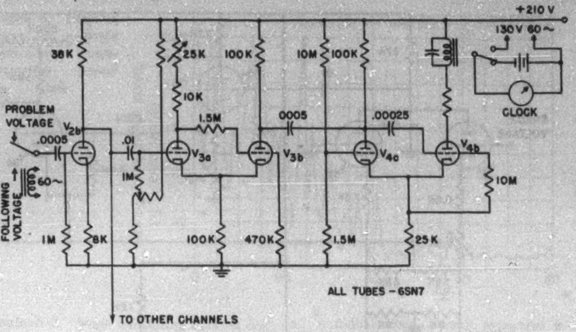


Figure 4. Diagram of time-on-target circuit

The Time-On-Target Circuit

Since time-on-target scores have been used frequently by other investigators in this field it was decided to incorporate scoring of this type in the electronic pursuit apparatus. The circuit of Figure 4 was devised to operate a clock whenever the absolute value of the error voltage was less than certain preset values, say 5, 10, and 20 per cent of the maximum problem voltage. A description of one of the three channels of time-on-target scoring follows.

The problem and manual control voltages S_p and S_c are alternately connected to a differentiating circuit by the vibrating relay thus producing a series of alternately positive and negative pulses at the grid of amplifier tube V_{2b}. The height of these pulses is proportional to the difference between the two input voltages. After amplification, these pulses are fed into three channels, one of which is shown.

The resistance network in the grid circuit of V_{3a} permits adjustment of the d-c level upon which the pulses are superimposed and consequently the size of error which will operate the clock relay. The circuit of V_3 is an amplitude discriminator, also known as a slicer or Schmitt trigger circuit, consisting of a direct coupled amplifier with positive feedback adjusted to the verge of instability. In its quiescent state, V_{3a} is cut off and V_{3b} is conducting. Whenever one of the positive pulses at the grid of V_{3a} is of sufficient height to cause conduction, the situation quickly reverses due to the positive feedback and V_{3b} is cut off thus producing a sudden rise of plate potential at V_{3b} which is essentially independent of the incoming pulse height so long as this pulse exceeds the preset critical value. The circuit returns to its original quiescent state when the tail of the input pulse drops below a critical voltage essentially equal to that at which the first switching action took place.

Tube V_4 and its associated circuit are a monostable multivibrator, also known as delay or single shot multivibrator. In the absence of any trigger pulses V_{4b} conducts steadily thus closing the relay in its plate circuit which in turn causes the clock to run. However, if the error signal is sufficiently large, trigger pulses will be supplied from the amplitude discriminator circuit at the rate of 60 per second to tube V_4. The trailing edge of each trigger

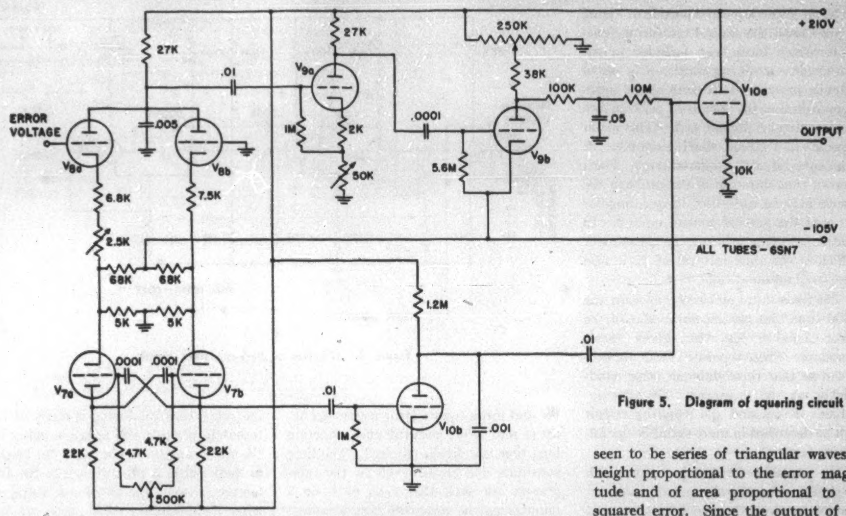


Figure 5. Diagram of squaring circuit

pulse causes the monostable multivibrator to operate, thus cutting off the current in V_{4b} for about one-fortieth of a second. The average current through the clock relay is then reduced to about one-third of its quiescent value thereby allowing the relay to open.

By using higher than rated voltage to operate the clock motor it was found on a number of checks that synchronous speed was attained in negligible time. Also by applying a direct voltage to the motor winding when the relay opens the motor is stopped abruptly thus giving a more accurate time indication.

The Squaring Circuit

As another measure of performance it was thought desirable to determine the mean square error of the humans' performance, since the measure would be analogous to certain statistical results used in other psychological experiments. The problem of measuring mean square error becomes one of devising a circuit to accept the error signal voltage, which may be of either polarity, and from this signal generate a voltage proportional to the squared error which is suitable for application to a conventional Miller integrator. This was accomplished as follows:

In Figure 5 the tube V_7 with its associated circuit is a free running, positive bias multivibrator with a frequency of about 2,000 cycles per second. Tube V_8 is two degenerative amplifiers with a common output and with cathodes connected to the cathodes of the multivibrator such that the amplifiers are alternately cut off at the multivibrator frequency. The combined output of tube V_8 contains a square wave whose amplitude is proportional to the error voltage.

This square wave is amplified in V_{9a} and then rectified in the grid circuit of V_{9b} producing at the plate of V_{9b} a voltage component proportional to the error signal in magnitude, but always of positive polarity. A simple resistance capacitance filter then removes the 2,000 cycle component from this voltage before it is applied to the grid circuit of V_{10a}.

Meanwhile, the square wave voltage from the 2,000 cycle multivibrator is fed to the grid of V_{10b}, a keying tube, which generates a saw-tooth wave at its plate due to the slow charging and rapid discharging of a small capacitor. This saw-tooth wave is then superimposed on the always positive error voltage at the grid of V_{10a}. With zero error voltage delivered to the circuit, the most positive excursions of combined saw-tooth and direct voltage at the grid of V_{10a} just barely reach the cut off value of the tube. When an error voltage does exist, the total voltage at the grid of V_{10a} will become more positive in direct proportion to the magnitude of the error voltage thus permitting a proportional height of the saw-tooth wave to be reproduced at the output terminal. The output voltage is then seen to be series of triangular waves of height proportional to the error magnitude and of area proportional to the squared error. Since the output of the squarer is fed directly to an integrator there is no need to filter out the 2,000-cycle-per-second frequency and its harmonics.

The four potentiometers shown provide means of adjustment for zero coincidence, equal sensitivity to positive and negative errors, over-all sensitivity and d-c level at the grid of V_{10a}.

A Brief Discussion of Some Studies Being Made by Means of the Apparatus

One of the more difficult problems encountered in the study of the behavior of human beings is the separation of the effect of various changes in the problem from the effect of variability from time to time in particular individuals and the variability between different individuals. This is a statistical problem which requires the accumulation and interpretation of large quantities of data. The problem especially calls for experiments in which several variables are studied in numerous combinations so that the interactions between variables can be determined as well as the primary effects of each factor.

Since there are a number of variables available for study on the apparatus, several exploratory studies have been made to determine the general effect of changing some of them. A typical series of studies was set up as follows:

1. The problem generator was set to produce a sine wave signal.

2. Two frequencies of the sine wave signal were used, about 18 cycles per minute and 35 cycles per minute.

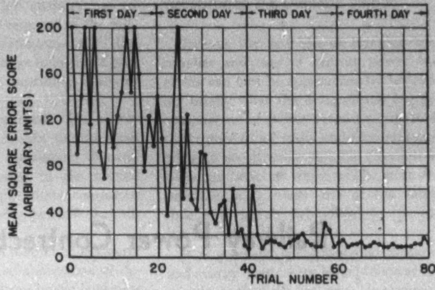
3. Two types of manual control were used, one involving a wrist movement and the other a movement of the entire forearm and hand.

4. The manual control scale factor was varied over a range of about six to one. On the arm control this involved an angular range of movement from 11 degrees of arc to 63 degrees of arc.

5. The visual display scale factor was varied in several steps involving total display amplitudes of one-fourth to four inches.

6. Both following and compensatory pursuit types of problems were used.

Figure 6. Learning curve for a single individual on a compensatory pursuit task



The mean square error (MSE) score was used as the figure of merit in these studies, a low MSE score meaning a better performance or an easier problem than a high MSE score. Some general and tentative conclusions formed on the basis of these preliminary studies are as follows:

1. Problems of the following pursuit type are much easier than compensatory pursuit type of problems.

2. Better scores are obtained at the low, sine wave frequency.

3. There is little if any difference between arm and wrist controls.

4. Manual control movements of large angular motion result in better performance than movements of small angular arc, especially at low frequencies.

5. There is a systematic improvement in performance as the amplitude of motion of the display is increased from one-fourth to 4 inches.

These results are in general agreement with previous experience and while they do not present any new or startling evidence they have served as a check on the over-all suitability of the electronic pursuit apparatus. Thus far no attempt has been made to determine precisely what interaction effects there may be between these different variables, for example, whether certain combinations give especially high or low scores.

A matter of considerable importance in the study of human behavior is the question of learning. This question is of importance in several ways. For example, a statistical study of the effect of a variable such as display scale factor may require that all subjects have reached their ultimate level of learning. In other cases it may be of interest to study the effect of the several variables on the initial performance of naive individuals or on the rate of learning. It also is of importance to know whether the combination of variables that gives the best initial performance is the one that results in highest ultimate skill. Some interesting data on the general question of learning have been collected.

Figure 6 shows a learning curve for a single subject working on a problem with the following characteristics.

1. Sine wave problem, frequency 29 cycles per minute.

2. Compensatory pursuit task.

3. Display amplitude, 4 inches.

4. Arm type of manual control with 11 degrees angular arc of movement.

This subject was tested on four nonconsecutive days, twenty one-minute trials per day. The curve on Figure 6 is a plot of MSE score in arbitrary units versus trial number. It is interesting to note the long term trend of the subjects' learning evidenced by the decrease in MSE score as well as the decreasing fluctuation of the score in any one day's trials. The obvious conclusion to be drawn from these data is that learning is a factor of prime importance in some pursuit tasks.

A number of other studies are either partially completed or in the planning stage. A brief description of some of these studies follows:

1. Graphic records of error, manual control movement, and acceleration of the manual control movement are being taken. These records will be analyzed in terms of several characteristics, such as maximum acceleration, acceleration patterns, and number of discrete corrective movements per cycle or per second. Plans are underway to analyze these records by correlation techniques to separate the random effects from the systematic effects.

2. A study of the effect of reaction time in continuous following tasks is in progress. This study is directly concerned with the problem of determining a suitable transfer function for the human.

3. Manual controls with proprioceptive feedback are under construction. The first control of this kind will have the properties of inertia, springiness, and velocity damping in adjustable and calibrated amounts.

4. A 2-dimensional pursuit apparatus is now under construction.

5. The effect of rate control, aided rate control and other modifications of the control path will be studied.

One of the major objectives of these studies is the determination of the range of problem variables to which the human can learn to adapt himself and perform at a maximum level of skill and in a stable manner.

References

1. Some Engineering Aspects of the Mechanism of Body Control, **Robert Mayne.** *Electrical Engineering*, volume 70, March 1951, pages 207–12.

2. Certain Characteristics of the Human Servo, **F. V. Taylor.** *Electrical Engineering*, volume 68, March 1949, page 235.

3. Some Characteristics of a Human Operator, **J. A. V. Bates.** *Journal*, Institution of Electrical Engineering (London, England), volume 94, 1947, pages 298–304.

4. Theory of the Human Operator in Control Systems, I and II, **K. J. W. Craik.** *British Journal of Psychology* (London, England), volume 38, 1947, pages 56–61 and 142–48.

5. The Application of Operational Analysis to Human Behavior, **D. G. Ellson.** *Psychological Review*, American Psychological Association (Washington, D. C.), volume 56, 1949, pages 9–17.

6. The Discontinuous Functioning of the Human Operator in Pursuit Tasks, **W. E. Hicks.** *Quarterly Journal of Experimental Psychology*, American Psychological Association, Washington, D. C., volume 1, 1948, pages 36–51.

7. The Nature of the Operator's Response in Manual Control and Its Implications for Controller Design, **A. Tustin.** *Journal*, Institution of Electrical Engineering (London, England), volume 94, 1947, pages 190–202.

8. Handbook of Experimental Psychology (book), **S. S. Stevens.** John Wiley and Sons, Inc., New York, N. Y., chapter 35, **P. M. Fitts**, 1951.

9. Electronic Instruments (book), **Greenwood, Holdan, MacRae.** McGraw-Hill Book Company, New York, N. Y., Radiation Laboratory Series, volume 21, 1947.

10. Waveforms (book), **Chance, Highes, MacNichol, Sayre, Williams.** McGraw-Hill Book Company, New York, N. Y., Radiation Laboratory Series, volume 19, 1947.

No Discussion

Railway Power Contracts

T. M. C. MARTIN
ASSOCIATE MEMBER AIEE

RAILWAY power contracts should be based upon rate schedules and general provisions that harmonize the respective interests of the power supplier and user. Elimination of the customary demand charge and the substitution therefor of a straight energy charge supplemented with a minimum bill provision appears to offer considerable assistance toward this goal. By this means railway managements would be freed from unnecessarily restrictive deterrents to good operation, and utility companies would at the same time be assured of earning sufficient revenue to cover fully all operating costs and a return on the capital invested in railway power facilities commensurate with the risks involved and with the return from capital devoted to other industrial power services.

Previous Papers

Two papers, one by Philip Torchio, the other by E. R. Hill, appeared some years ago,[1,2] that dealt particularly with railway power supply and the fundamentals of power contracts covering service to railway loads by public utility companies. Among the questions raised in these papers were the following:

1. Generation of power by railroad or utility company.
2. Character of service by utility—3 phase, 60 cycle or other—basic power contract.
3. Ownership of and charges for special conversion equipment.
4. Basis of power charges, demand and energy components, et cetera.

These may be considered in order:

Generation of Power by Railway or Utility

This question has been pretty well answered in the intervening years. Railroads appear to be no longer interested in building their own power plants. In one or two instances it would appear that there might be an economic basis for a railroad building a modern generating station sufficient to meet its own peaking requirements through sale of surplus secondary or dump power direct to other industrial users along its tracks or to electric utilities. This is not the prevailing situation, however.

Character of Utility Service

It seems necessary to recognize that the public utility industry has become tooled as it were for the production of but one product, insofar as large industrial loads are concerned, and that is 3-phase 60-cycle alternating current. The industry also supplies residential customers with single-phase service, but it is widely recognized that this is only feasible where the numbers of such users are relatively large and their individual requirements correspondingly small thus affording the utility great time diversity and the opportunity to economically assemble these small loads into fairly large balanced 3-phase blocks for imposition on its major transformation, transmission, and generation facilities.

Special Conversion Equipment

If the electrification system of the railroad requires the use of direct current, or current alternating at some frequency other than 60 cycles, or the purchase of current at single phase, then the necessary current conversion, frequency changing, or phase balancing as the case may be, probably will be treated best as a separate matter at least as far as the basic power contract is concerned. Acceptance of this principle need not mean that the railroad is forced to perform itself such alteration of basic 3-phase 60-cycle current. In each instance a railroad would have an opportunity to choose between purchasing, installing, and operating the necessary conversion equipment or separately contracting with the utility to furnish and operate such equipment. The question may resolve itself into one of comparative economics.

If it is agreed that the processing of power beyond the 3-phase 60-cycle stage is properly considered an ancillary problem, both physically and contractually, it will be possible to proceed to the question of the basic power contract and discuss the latter from the viewpoint of both the railroad and the power supplier.

Basic Power Contract

In discussing basic railway power contracts it will be desirable to limit the scope of the discussion somewhat by excluding certain types of railway operations which impose special problems upon power suppliers and require separate study and evaluation. Such exclusions would include all urban mass passenger transportation mediums such as subways and elevateds. Also excluded would be main-line railroads that electrify only their suburban services leaving their long-haul trains to be handled with other types of motive power. The reason for eliminating from this inquiry these particular kinds of operations is owing to the special peaking effects associated with movements composed entirely of commuter services. This restriction need not be extended to include main-line railways that are completely electrified even though a portion of their normal traffic components may include commuter services of a substantial character in certain areas. As long as all services both freight and passenger are handled with electric motive

---

Paper 52-18, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted October 19, 1951; made available for printing November 21, 1951.

T. M. C. MARTIN is with the Bonneville Power Administration, Portland, Oreg.

power the analysis that follows will be appropriate.

Railways in the category under discussion, which incidentally, includes most of the class *I* line-haul carriers in the United States, are essentially bulk freight haulers with relatively small portions of their total train movements devoted to passenger service of any kind, long-haul or commuter. Careful analyses of thousands of train sheets by numerous competent investigators have served to develop the fact that where freight traffic forms the bulk of the ton-miles handled there is a very high degree of inherent time diversity.

Hypothetical load curves based on such traffic have revealed a very considerable variance in the occurrence of daily peaks. There has been no observed tendency for these daily railway peaks to fall into any invariable pattern, nor and what is more important, has there been any indication that the railway peaks exactly or nearly coincide with the daily system peak of the remainder of the utility's electrical customers.

Mr. Torchio remarked on this point in his paper[1] "Furthermore, although the railway's load factor is, in general less favorable than the utility's, when the two loads are superimposed with their diversity in time of respective maximum demand, the combined load factor is improved and conditions are created for realizing substantial economies in plant investment."

Demand and Energy Factors

The conventional industrial power contract is centered in two principal rate factors:

1. A demand factor, and
2. An energy factor.

Specified proportions of the total charges are ascribed to each element. Contractual arrangements embodying these elements have been fairly successful in bringing into harmony the separate interests of the power supplier and power user in industries other than railroads. In the case of railways, however, the negotiation of a mutually satisfactory power contract has never been too harmonious a process, and the resulting documents have not infrequently been a continuing source of irritation to both parties. Mr. Hill[2] in speaking of the desirable attributes of a good railway power contract said in part: "Power contracts should be as simple as possible as to form and should eliminate all unnecessary conditions and stipulations that might lead to misunderstandings and disputes. No guarantee as to load factor should be required of the railroad, nor should guarantees as to the minimum demand charge following a previously established maximum be required."

Mr. Torchio also warned, "The operation of a railroad is subject to occasional large variations of maximum demand from causes extraneous to the normal operation. These fortuitous conditions should not be allowed to penalize the railroad unduly."

What Messrs. Hill and Torchio clearly recognized is that the application of demand charges in railway power contracts is a matter that solicits the utmost discretion and understanding on the part of the power supplier. Unlike most other industrial power customers, railroads cannot manipulate their demands very much. The freight traffic, which will largely determine the magnitude of power demands, is offered to them as common carriers by other parties in accordance with prevailing general business conditions and must be handled nearly as soon as offered.

That is the essence of the transportation business. The total energy requirements will vary almost directly with the total number of ton-miles moved, but the rate at which energy is taken, the demand, will be determined almost entirely by the speed at which trains are moved and the times when they reach the ruling grades. There is no set pattern, obviously, for the latter.

Measurement of Demand

Even the determination of what constitutes demand requires evaluation. It has become a fairly common practice now in the utility business to employ what is termed a 30-minute integrated basis for demand. This is accomplished by measuring the kilowatt hours per clock half-hour. Multiplied by two these are recorded as kilowatts of demand. Some railway power contracts have been based more liberally on clock-hours rather than half-hours. On the other hand, some railway power contracts have been based on demand intervals as short as five minutes which approaches an instantaneous basis. Another restrictive feature known as a ratcheted demand has been incorporated in some railway power contracts. This feature may provide that a demand established in one month be used as a minimum demand for billing purposes in future months for perhaps a year or, in some instances, for the remainder of the term of the contract.

Substitution for Demand Factor

Perhaps the problem of bringing the supplier and user of railway power together can best be approached by reviewing briefly what the demand and energy components of power rates are expected to accomplish normally and then endeavoring to see if some suitable alternative basis can be found which will be more acceptable to railroads and at the same time not entirely prejudicial to the interest of the utility.

It is generally contended that demand and energy charges are levied to cover respectively the fixed and variable elements of the cost of supplying power. Simply stated the demand charge is to cover the capital costs, including interest, depreciation, and taxes, of the entire special facilities devoted exclusively to railway use as well as some allocated portion of the capital costs of joint facilities. The energy charge is to reimburse the utility for materials and labor, such as fuel and ordinary maintenance, used up in daily operations. The former relates to costs which accrue whether a single ton moves over the railway or not, whereas the latter is applicable to costs which vary directly with the volume of traffic and therefore with the amount of electric energy consumed.

There are at least two ways in which this objective can be approached without resort to the demand charge. One formula that has been successfully employed for a great many years in one particular railway power contract is a fixed charge of so many dollars per route mile per month. This, supplemented with an energy component, has constituted the basis for power charges. On the whole the arrangement has worked very well. It has freed the railway management from the restrictive influence of ordinary demand charges and at the same time has yielded revenue to the power supplier in amounts sufficient to have earned two rate reductions in the past ten years. Another possible method of determining the amount of the fixed charges is through the use of a minimum bill.

Either of the above methods accomplishes the objective of placing a floor under load factor and assures the power supplier that the full capital costs of all special investments in facilities for the exclusive use of the railroad will be recovered irrespective of the volume of railway traffic.

The minimum bill appears more practicable. It can be expressed in simpler terms and tied to a more realistic and stable parameter than route mileage.

The minimum bill can best be applied in the form of a straight energy charge with a requirement that not less than a specified amount of money be paid each month.

Conceivably, this minimum could be separately determined for each project. It might be easier, however, for a utility to compute average costs and establish a rate schedule which relates these costs to some readily available measure such as total capacity.

Suggested Rate Schedule

1. Energy sold under this schedule shall be paid for at the rate of .... mills per kilowatt-hour of net energy delivered, et cetera.

2. The net minimum charge under this schedule shall be .... cents per month times the aggregate kilovolt-ampere capacity of all points of delivery.

In this discussion all reference to power factor correction, phase balancing, et cetera, has been eliminated since these matters as well as numerous other more or less stock provisions of industrial power contracts have never been too controversial.

References

1. Railway Power Supply, Philip Torchio. *Electrical Engineering (AIEE Transactions)*, volume 59, September 1940, pages 550–53.

2. Power Supply for Railroad Electrification and Fundamentals of Power Contracts, E. R. Hill. *Electrical Engineering (AIEE Transactions)*, volume 59, September 1940, pages 560–62.

No Discussion

A New Electric Locomotive for the Pennsylvania Railroad

F. D. GOWANS
NONMEMBER AIEE

B. A. WIDELL
MEMBER AIEE

A. BREDENBERG
MEMBER AIEE

Synopsis: Simplicity and reliability are the keynotes of the design of the new single-phase a-c electric locomotives for the Pennsylvania Railroad. Basic components tried and proved by years of successful operation were deliberately chosen and combined into a new design. Full advantage was taken of the latest improvements in materials and manufacturing processes to create a motive power unit of high capacity which is economical to build, operate, and maintain.

RAPIDLY increasing use of diesel-electric locomotives in the past few years has caused much speculation on the future of railroad electrification. Proponents of the diesel-electric point to its lower first cost, while the electric locomotive enthusiast claims lower maintenance and operating cost. The final answer will be an economic one, dependent upon such things as future electric power developments and new sources of fuel oil supplies.

The important fact is that the railroads have accepted the electric drive as the best and most efficient means yet developed for propelling railroad vehicles. Now it is of prime importance that the components of this drive, whether for straight electric or diesel-electric locomotives, and the vehicle which it propels be continually improved. The builder should make every effort to include in his new designs improvements in materials and processes, together with any other factors that will contribute to longer life, higher availability, and lower over-all cost.

What Type Locomotive?

One important factor in the success of a locomotive is the type and arrangement of the mechanical parts. Past designs with many axles (some idle), complicated trucks, and crowded cabs are good examples of high first cost, difficult manufacture, and expensive maintenance. When the General Electric Company was afforded the opportunity of designing a new freight locomotive for the Pennsylvania Railroad, serious consideration was given to all phases of this problem.

Freight traffic on the road's electrified lines was studied over a period of years to determine what size unit would give the best utilization. A 4-axle locomotive unit capable of operating in 1-, 2-, 3-, or 4-unit combinations was chosen. Experience with existing electric freight locomotives showed that tonnage ratings should be based on average tractive effort over the ruling grades equal to 16 per cent adhesion. At the 60,000-pound axle loading limit specified by the railroad, this amounted to a 1-hour tractive effort rating of 9,600 pounds per axle, or 38,400 pounds per unit. Table I shows how such units could handle the Railroad's freight traffic through a typical industrial cycle. This size also is most desirable because of its simple mechanical construction and the additional availability it offers over larger units.

Selection of Electric Equipment

Equally important is the type and arrangement of the electric equipment. The locomotive type will, of course, largely determine this. The Railroad was willing to consider motive power of types other than its present straight a-c locomotive, if advantages were indicated.

The a-c commutator motor is inherently a high-horsepower machine. Its chief advantage lies in its ability to take 25-cycle power from the trolley without intermediate conversion equipment. For this reason locomotive ratings are based entirely on the traction motors. With overvoltage transformer taps, motors can be operated at extremely high short-time ratings for accelerating heavy trains and ascending grades. The relative simplicity of the equipment permits good arrangement, making ample space available for maintenance and inspection. The additional weight and cost of the a-c motor are offset by the conversion equipment necessary if d-c motors are used.

Advantages were decidedly in favor of continuing with the straight a-c type locomotive. This has been the basic motive power on the Pennsylvania's electrified lines since 1931 and has given commendable service. Railroad schedules, operating practices, and shop facilities have been built up around it. Twenty years' experience with this plant has demonstrated that simplicity of mechanical construction and of electric circuits is the key to low maintenance and reliable operation.

Characteristics

Table II gives weights, dimensions, and ratings of a single unit. Units may be coupled, see Figure 1, for multiple operation from either end, with as many as four units operating together. So far, only *A* units have been built, but the design is

Table I

Number of Units	Per Cent of Freight Traffic Peak	 Normal	 Depression
1	13	34	64
2	60	63	36
3	22	3	0
4	5	0	0

Paper 52-23, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted October 23, 1951; made available for printing December 13, 1951.

F. D. GOWANS, B. A. WIDELL, and A. BREDENBERG are with the General Electric Company, Erie, Pa.

Table II. Locomotive Unit Data

Classification *B-B-240/240-4* GEA 632
11,000 Volts, 25-Cycle Alternating Current

Weight	
Total locomotive, fully loaded	240,000 Lb.
Per driving axle, fully loaded	60,000 Lb.
Dimensions, approximate	
Track gauge	56½ In.
Length inside knuckles	54 Ft 3 In.
Height over cab roof	14 Ft 0 In.
Height, trolley locked down	15 Ft 0 In.
Width over cab sheets	10 Ft 0⅛ In.
Width over-all	10 Ft 7 In.
Total wheel base	39 Ft 10 In.
Rigid wheel base	11 Ft 0 In.
Length between centerplates	28 Ft 10 In.
Wheel diameter	48 In.
Coupler height	34 In.
Clearance, motor gear case to rail	4⅝ In.
Minimum curve	275 Ft radius, 21 Deg.
Ratings	
Tractive effort at 25 per cent adhesion	60,000 Lb.
Tractive effort, continuous rating	35,400 Lb.
Tractive effort, 1-hour rating	38,400 Lb.
Speed, 1-hour rating	25.5 Mph.
Maximum speed	65 Mph.

readily adaptable to *B* units, considered more desirable when more than two units are coupled together.

Characteristic speed-tractive effort curves for motoring are shown in Figure 2. These are for a 2-unit 5,000-horsepower locomotive.

In recent years rheostatic braking, now commonly used on diesel-electric locomotives, has established its worth in dollars and cents. This usefulness, especially in freight service, has now resulted in its application on a straight a-c locomotive for the first time in the United States. Braking characteristic curves for the 2-unit locomotive are shown in Figure 3. The a-c commutator motors are well adapted for this use, rating 750 horsepower each during braking within the limits of maximum excitation and commutation.

Mechanical Design

The mechanical design, with a streamlined, single-end cab carried on two 2-axle swing-bolster trucks, is quite different from earlier a-c freight locomotives. Except for the pantographs on the roof, this locomotive resembles the conventional diesel-electric *A* unit. Collision framing at each end is of sturdy construction. The streamlined nose affords additional protection and excellent visibility for the engine crew.

Doors are located at each end of the unit, on each side of the operating compartment, and on the right side of the apparatus compartment toward the rear. Framed grilles in each side of the apparatus compartment as high as possible from the ground, permit entrance of clean cooling air for the traction motors, controls, and transformer.

The entire cab and platform structure is fabricated from steel sheets, plates and shapes; and is designed for subassembly of parts and equipment groups. Arrangement of apparatus in the cab provides maintenance aisles on each side, with doors into the operating compartment at the front and access to the rear end door. Hatch covers over all equipment permit complete accessibility from the roof.

Locomotive platforms were formerly designed to carry the entire equipment load and withstand the buff without undue deflection. The cab side sheets and roof were part of the platform load. No doubt they added some stiffness to the structure, although this was not considered in the calculations. By designing the side sheets to prevent buckling, they can be used as side trusses to carry the entire load. The consequent reduction of material in the platform amounts to approximately 40 per cent, as compared with a five per cent increase in side sheet material. This construction, see Figure 4, results in considerable weight and cost reduction and has been adopted for the new locomotive design.

The new assembly routine required by this design is well suited to the practice of subassembly. Platforms are set up on the assembly floor and supported on their centerplates and at intermediate positions as required to give the upward camber. Tests were conducted to determine the correct proportion of stress in each member and the camber required to obtain it. Subassemblies of equipment are installed on the platform and interconnected. After this, the front portion of the cab, including the nose and operating compartment, is put in place and welded to the platform and collision structure. The rear portion, including side sheets and roof, is then set in place and welded to the front portion and the platform. When the intermediate supports are removed after welding, loading of the side sheets is indicated by deflection of the platform to its normal position.

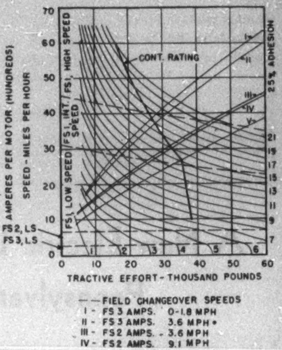


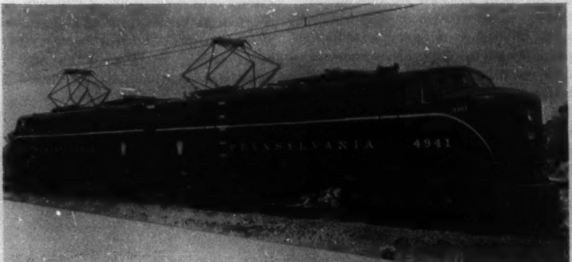
Figure 2. Speed-tractive effort curves for one 4-motor locomotive unit with 83/21 gearing and 48-inch wheels

Conventional 2-axle, swing-bolster trucks are used, see Figure 5. These have cast-steel frames, drop-center equalizers, and clasp brakes. The spring rigging is generally similar to that used on Pennsylvania Railroad passenger cars. Coiled springs support the truck frame on the equalizers, and an arrangement of coil and elliptic springs supports the bolster on the spring plank.

Rolled steel wheels, 48 inches in diameter, are mounted on forged carbon-steel axles, which support the truck on 7- by 12-inch antifriction bearings. Pedestals, bolsters, centerplates, and side bearings are fitted with renewable carbon-steel liners. Case-hardened pins and Railite bushings are used throughout, except on the bolster swing hangers which have case-hardened bushings.

Rubber draft gear is used at both ends. The front end is fitted with type *E* coupler

Figure 1. Two units operating in multiple as a 5,000-horsepower locomotive



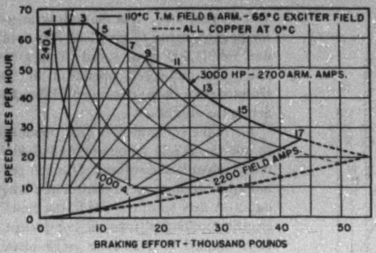


Figure 3. Speed-braking effort curves for one 4-motor locomotive unit with 83/21 gearing and 48-inch wheels

and coupler enclosure, while the rear end has a tightlock coupler.

Location of Apparatus

Two factors largely determine locomotive inspection and maintenance costs: reliability of the equipment and its accessibility in the locomotive. Although the latter may be secondary, neither can be ignored if the locomotive is to be considered satisfactory. In designing these units, prime consideration was given to the location of apparatus. In fact, this was planned first and, with few exceptions, the mechanical design was made to suit.

Figure 6 shows that the traction-motor control groups and blowers are located directly over the trucks. Bus bar and cable runs are thereby kept to a minimum and blowers discharge directly downward into the motors. Conditions at the rear end are ideal, but at the front end the control group and one blower had to be offset to accommodate the operating compartment. Both motor control groups receive power from the transformer and its control group which, together with the auxiliary equipment, is located centrally. Motor controls are accessible through doors along each side aisle. A crosswise maintenance aisle for the transformer group extends from one side aisle to the other.

The location of the transformer high-voltage bushing and Thyrite resistance lightning arrester on the roof is unusual. The transformer tank is the full height of the apparatus cab, allowing the bushing to protrude through a cutout in the hatch cover directly under the front pantograph. An insulated bus on the roof connects both pantographs, the lightning arrester, and the transformer. It also serves as the terminal for the ground switch. This arrangement eliminates all 11,000-volt cable, cable preparation, and conduit, together with the shields and barriers necessary for the protection of workmen inside the apparatus cab from contact with this voltage.

The braking resistor and braking exciter are located at the rear of the cab.

Figure 4 (left). Side sheet contruction used to form supporting truss for platform

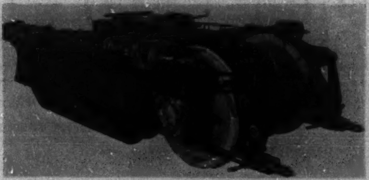


Figure 5 (below). Two-axle, swivel truck of conventional design used on this locomotive

Air reservoirs are located under the platform between the trucks.

Equipment and controls in the operating compartment are grouped conveniently for the engine crew. The rear wall contains cabinets for the field-shunting control, and a clothes locker. Compartments for supplies and tools are arranged under the raised platform on both sides.

Air brake equipment, train signals, and small control pieces are located in the nose. Here they are accessible and convenient to connections from the engineman's position.

The two pantographs on the roof of each unit are accessible from inside the cab by means of a ladder and hinged door in the hatch cover. This door is mechanically interlocked with the pantograph grounding switches to prevent access to the roof when the pantographs are ungrounded.

Sand boxes, fire extinguishers, and other miscellaneous equipment are arranged in locations suitable for their particular functions. Each locomotive unit has a sandbox capacity of 20 cubic feet.

Auxiliary Equipment

Considerable improvement has been made in auxiliary equipment. For instance, the use of axial-flow blowers has contributed to better arrangement and accessibility of equipment because they occupy about half the space of centrifugal blowers of the same rating. The blowers incorporate an air-cleaning feature which removes about 60 per cent of the dirt, water, and snow from the air furnished to the electric equipment. Since they are driven by 3-phase induction motors, the expense of brushes, commutator maintenance, and freight inspection is eliminated. Because of their exceptionally good characteristics for ventilating this type of equipment, these blowers have been used throughout the locomotive.

An auxiliary motor-generator set provides 3-phase power for the blowers and d-c power for control and battery charging. It consists of a 2-speed 25-cycle single-phase induction motor driving a 125-cycle 3-phase alternator and a 37.5-

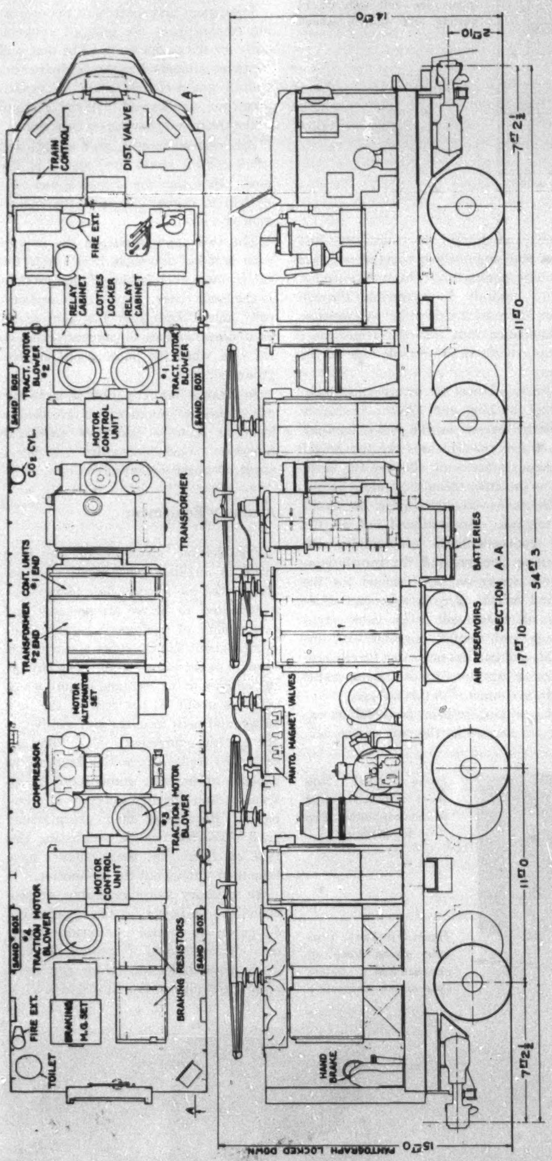

TRAIN CONTROL
DIST. VALVE
FIRE EXT.
RELAY CABINET
CLOTHES LOCKER
RELAY CABINET
TRACT. MOTOR BLOWER #2
TRACT. MOTOR #1 BLOWER
SAND BOX
SAND BOX
MOTOR CONTROL UNIT
CO2 CYL.
TRANSFORMER
TRANSFORMER CONT. UNITS #1 END
#2 END
MOTOR ALTERNATOR SET
COMPRESSOR
#3 TRACTION MOTOR BLOWER
MOTOR CONTROL UNIT
#4 TRACTION MOTOR BLOWER
SAND BOX
SAND BOX
BRAKING RESISTORS
FIRE EXT.
BRAKING M.G. SET
TOILET
A
A
14'-0"
2'-10"
PANTO. MAGNET VALVES
HAND BRAKE
AIR RESERVOIRS
BATTERIES
SECTION A-A
7'-2½"
11'-0"
17'-10"
54'-3"
11'-0"
7'-2½"
15'-0" PANTOGRAPH LOCKED DOWN

Figure 6. Location of principal pieces of apparatus in the locomotive cab

volt d-c generator. The motor is arranged for pole changing to give half (750 rpm) and full (1,500 rpm) speed. This gives half- and full-speed operation of the blowers since they are connected directly across the alternator.

Air brake equipment is *8EL*, modified to use the *DS-24* automatic and *S-40-F* independent brake valves. This will permit application of speed control at a later date. A new feature is use of the diesel-electric locomotive type air compressor. It is coupled to a 25-cycle single-phase induction motor and arranged for unloaded starting. It has a capacity of 220 cubic feet per minute at 725 rpm.

Two main air reservoirs, having a combined volume of 50,000 cubic inches, are located under the platform. The air coolers have ample capacity. The entire system has the very desirable feature of being drained into the main reservoirs.

Brake rigging on the truck has been arranged with one cylinder per wheel to provide a 75 per cent braking ratio with 50-pounds-per-square-inch cylinder pressure.

Traction motor excitation, during rheostatic braking, is provided by a 1,500-rpm 25-cycle single-phase induction motor driving a 10-pole 10-volt d-c generator.

Traction Motors

The *GEA 632* motor, see Figure 7, is a 16-pole 25-cycle single-phase, commutator-type a-c motor with commutating poles and pole face windings. It is rated 625 horsepower continuously at 230 volts. The maximum operating speed is 1,800 rpm, corresponding to a speed of 65 miles per hour on these particular locomotives. The motor is mounted on the axle and a nose support on the truck transom. A cushion gear, pressed on the axle, is driven by the motor pinion.

It has long been recognized that better performance can be obtained from a straight-electric locomotive traction motor than from a diesel-electric or motor-generator locomotive motor. On a-c locomotives this extra performance is obtained by providing overvoltage taps on the transformer. The multipole low-flux per-pole type of single-phase, a-c motor is ideally suited to transmit this additional power. Although the design is largely predicated on considerations of standstill starting phenomena, it is also the best to withstand applications of overvoltage. The first motors of this general type were placed in service on the Pennsylvania Railroad in 1931. A very complete analysis of their design has been published.[1]

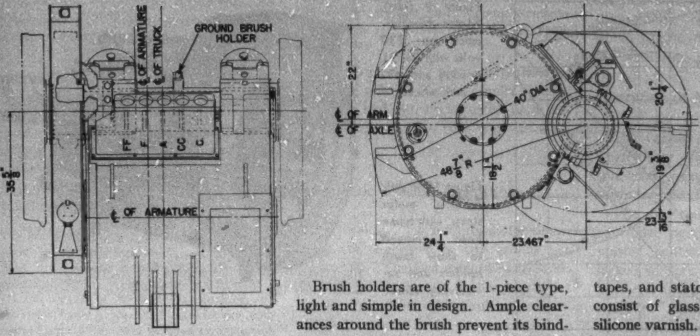


Figure 7. Outline of traction motor with principal dimensions

The present motor reveals the significant design progress made since 1931. The continuous horsepower per pound increased 44 per cent in the 1934 design, 13 per cent in the 1938 design, and 26 per cent in the 1951 design. Also, of interest mainly to motor designers, the comparisons of the speed-ratio power per pound index shown in Figure 8 serve to graphically indicate this progress. This index has been used before in presenting design progress trends.[2] It is basically proportional to:

$$\frac{\text{Rated torque} \times \text{maximum permissible speed}}{\text{motor weight}}$$

The index ratings shown cover the entire speed range in order to present a complete picture of the development of this motor. Each of the motors shown has approximately the same amount of ventilation per horsepower.

The characteristic performance of this motor at 230 volts is shown by Figure 9.

Electrical and Mechanical Differences

Electrical design points differing from former motors of similar type are:

1. A slightly smaller exciting-pole air gap and a larger commutating-pole air gap.
2. Armature slots that are open instead of partially closed, to permit using preinsulated coils with their inherent advantages.
3. Solid armature bar in place of the folded bar.
4. More armature slots per pole and shallower slots, to partially offset the additional losses in the solid armature bars at high speeds.

Mechanical details differ in many respects from the older type of a-c locomotive motor. Wherever possible these conform to the improvements made in d-c motor construction during the last few years, see Figure 10.

Brush holders are of the 1-piece type, light and simple in design. Ample clearances around the brush prevent its binding in the holder as a result of accumulations of dust.

Armature coils are of semi-1-piece split-throw design with flattened leads for connecting to the commutator risers. Less than half as many back end brazes are required as formerly. The bars have a straight portion at the back end connections of the winding to give additional surface for binding purposes.

Armature slots are sufficiently shallow so that finger ring supports are ~~not~~ required at the ends of the armature. Reenforcement consisting of a few 1/16-inch thick punchings at the ends of the stacking, as used on d-c machines, prevents flaring. Adequate creepage between armature conductors and ground is more readily obtained than with the finger rings used on previous designs of a-c motors.

A revolving brush-holder yoke permits inspection and replacement of brushes. This is located to line up the brush holders in correct relation to the main poles by a key set accurately with fixtures at the factory.

The cushion gear is similar to those used in the past, with the exception that tube-form synthetic rubber joints replace the flat or spiral steel springs formerly used to dampen the 50-cycle torque impulses. These joints consist of an inner steel cylinder to which is bonded a tube of rubber. An outer steel cylinder, not bonded to the rubber, holds the tube in initial compression. Deflection characteristics with applied load are closely linear after the initial breakaway pressure has been applied.

Insulation

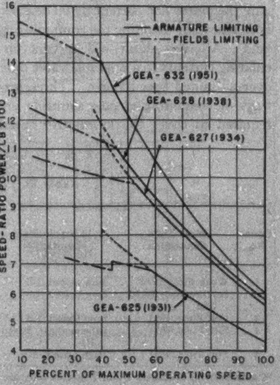
The continuous motor rating is given on the AIEE basis for class *B* insulation, 120-degree-centigrade rise on the armature and 130-degree-centigrade rise on the field as measured by resistance. Conductor insulation, armature polycoil tapes, and stator winding slot wrappers consist of glass and mica treated with silicone varnish.

Each motor is blown with 4,200 cubic feet of air per minute. Internal restrictions to air flow have been minimized to keep the blower power within economical limits.

Main Field Flux Control

At standstill, and during very low speed operation, circulating currents exist in the armature turns which are short-circuited by the brushes. These currents are induced by the 25-cycle main field flux. At low motor speeds they cannot be restricted by commutating-pole flux control. In order to hold these currents to a tolerable value, the main field flux must be limited in amplitude. This is accomplished by a combination of autotransformer and inductive shunt which weakens the field to 40 per cent of full strength (*FS-3*, Table III). As the locomotive speed increases, a speed voltage is generated in these short-circuited turns by an

Figure 8. Trends in continuous horsepower per pound of motor weight



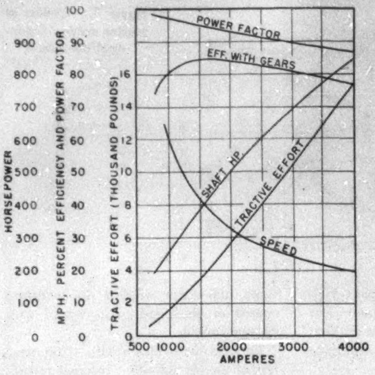


Figure 9 (left). Motor characteristics on 230 volts with 83/21 gearing and 48-inch wheels

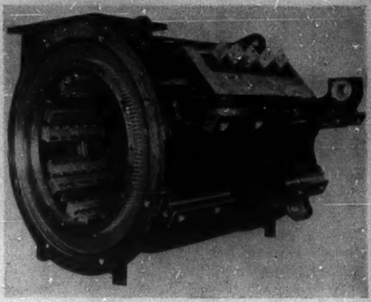


Figure 10 (right). Traction motor stator, with frame head removed to show brush holders and revolving yoke

out-of-phase component of the commutating-pole flux, often referred to as quadrature compensation. This voltage serves to reduce the circulating currents to the extent that, at 3 miles per hour, the autotransformer may be cut out. This gives 67 per cent main field strength (*FS-2*, Table III). As a result, for a given motor torque (locomotive tractive effort) the motor line current is significantly reduced. At 7 miles per hour, the circulating currents have decreased to the point where the main field shunting may be completely removed. Operation then continues in full field (*FS-1*, Table III) at higher speeds.

Commutating Field Flux Control

The quadrature compensation voltage mentioned previously is a function of both motor speed and the magnitude of the quadrature component of commutating pole flux. This component of flux, lagging the motor line current by 90 degrees, is produced by shunting the commutating field. A given shunt will restrict circulating currents to a tolerable value only over a limited speed range. Three settings of commutating pole shunting are required to cover the locomotive speed range. The low-speed shunt, a resistive impedance, is connected across the commutating poles from standstill to 26 miles per hour. As already noted, this is not fully effective below 7 miles per hour, and main field shunting must be resorted to (Connection numbers and 1, 2, and 3, Table III). Above 26 miles per hour commutation again becomes unacceptable with this shunt. Supplementary resistive-reactor impedances are connected over the intermediate speed range from 26 to 40 miles per hour (Connection Number 4, Table III). Additional impedance is connected in the high-speed range above 40 miles per hour (Connection Number 5, Table III).

The control is arranged to change from one speed setting to the next at the same degree of commutation. In addition to the adjustment for quadrature compensation, the in-phase compensation provided to commutate load currents, is set to the optimum commutation value at each of the speed settings. This value is the same as that required for d-c operation.

Design trends have been toward lighter motors which can be operated continuously in the lower range of locomotive speeds as well as at high speeds. This new motor has been designed with sufficient available quadrature compensation to provide a low best speed of commutation of 21 miles per hour, as compared to 25.5 miles per hour for older locomotives of similar capacity geared for the same maximum speed. Good performance is thus assured at relatively low speeds over long, heavy grades as well as in high-speed service where maximum horsepower output is desired.

Table III. Motor Connections

Conn. No.	Main Field Connection	Interpole Field Connection	Approximate Speed Range, Miles per Hour	Contactors Closed*
1	Weak Field, FS3	Low Speed, LS	0 to 3	F12, F13, F14R, F24R
2	Intermediate Field, FS2	Low Speed, LS	3 to 7	F11, F12, F14R, F24R
3	Full Field, FS1	Low Speed, LS	7 to 26	F11, F14R, F24R
4	Full Field, FS1	Intermediate Speed, IS	26 to 40	F11, F14R, F24R, F15X, F25X
5	Full Field, FS1	High Speed, HS	Above 40	F11, F14X, F24X, F15X, F25X

*Contactors for Numbers 1 and 2 motors only are listed.

Generator Operation

When the motors are operated as separately-excited d-c generators and loaded on resistors for rheostatic braking, 39 per cent of the line current must be shunted out of the interpole field for best commutation. To prevent changes in interpole winding temperatures from causing appreciable changes in the amount of shunted current, a portion of the permanent resistance used across the interpole field in a-c operation is connected in series with the interpole field. The remainder is used to shunt the required amount of current out of the interpole and resistor leg.

Control Equipment

Multiple-unit single-end control of the electropneumatic type is used. Two spring-raised air-lowered pantograph trolleys collect current at 11,000 volts from the trolley wire.

The main transformer (rated at 2,655 kva, 11,000/762 volts, 25 cycles, single phase) is Pyranol filled and forced ventilated.

Motoring Operation

The traction motors are connected in two parallel groups, each with two motors in series, see Figure 11. Varying voltage steps are provided by the well-known 3-preventive coil system of notching on the taps of the transformer secondary. A total of 21 control steps gives acceptable tractive effort increments, as indicated in Figure 2. The use of a motor-operated controller provides definite, simple interlocking between positions, and prevents a wrong sequence of operations.

The main and interpole field connections are under the control of a 5-position, motor-operated controller. This, in turn, is controlled by relays connected in the traction-motor circuit as follows:

1. Main field relay, of the phase-angle, motor type with contacts mounted on the rotor to give the 3-field connections *FS3*, *FS2*, and *FS1*. Its stator is excited from a 99-volt section of the main transformer, and its rotor from a potential transformer connected across the main fields of two traction motors connected in series. The rotor takes a position corresponding to the phase angle between the main transformer and traction-motor field voltages. This angle is relatively large at standstill and decreases as locomotive speed increases. Thus the relay is responsive to locomotive speed and can be used to control the traction-motor field connections.

2. Interpole field relays, one for each of the two interpole field connection changes, are identical except for speed setting. Each has an operating coil with one winding responsive to traction-motor armature voltage and the other, a differential winding, responsive to main field voltage. Thus the relay is operated by a combination of traction-motor voltage and current elements which gives the correct speed for interpole changeover.

Transformers and rectifiers permit using a d-c relay. This is simpler in construction and can be adjusted more accurately than an a-c relay for the same purpose.

Rheostatic Braking Operation

For braking, the traction-motor armatures are connected in two groups, see Figure 12. The two armatures in series in each group operate as d-c generators across a constant-value braking resistor. All four motor fields are connected in series across a low-voltage d-c exciter, the shunt field of which is supplied from the 37.5-volt d-c control circuit.

Control of braking effort and speed is obtained by varying the exciter shunt field. This is done manually through the braking handle on the master controller which operates only seven contactors in the exciter field circuit to give 17 braking steps, see Figure 3.

Each braking resistor is of the strip ribbon type, cooled by a d-c motor-blower set connected across it. With this connection, an increase in braking current causes a corresponding increase in air flow through the resistor, and power that would otherwise be wasted is used to drive the ventilating fans.

Operation of Auxiliaries

The various auxiliaries are driven by single-phase 25-cycle induction motors connected to taps on the main transformer secondary, see Figure 13.

The motor-alternator set is manually controlled by the selector handle on the master controller, mechanically interlocked so that the locomotive cannot be operated unless the handle is in a running position. The blowers are connected across the alternator and are automatically started by the closing of the alternator field contactor when the set comes up to speed. With the locomotive at standstill, the selector handle may be moved to give full- or half-speed operation of the set. In any case, full-speed operation is automatically obtained on the fifth controller notch. If necessary during severe snowstorms, the motor-alternator set, and thus the blowers, may be held at half speed by pinning the selector handle in the low-speed position.

Split-phase resistance starting is used on the motors of the auxiliary sets. A re-

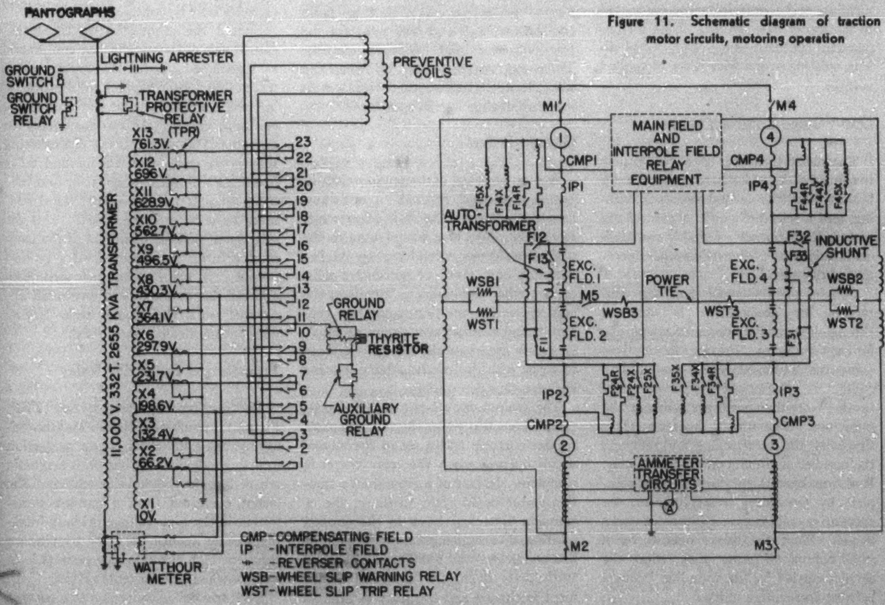

PANTOGRAPHS
LIGHTNING ARRESTER
GROUND SWITCH
GROUND SWITCH RELAY
TRANSFORMER PROTECTIVE RELAY (TPR)
PREVENTIVE COILS
X13 761.3V.
X12 696V.
X11 628.9V.
X10 562.7V.
X9 496.5V.
X8 430.3V.
X7 364.1V.
X6 297.9V.
X5 231.7V.
X4 198.6V.
X3 132.4V.
X2 66.2V.
X1 0V.
11,000 V. 332T. 2655 KVA TRANSFORMER
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
GROUND RELAY
THYRITE RESISTOR
AUXILIARY GROUND RELAY
WATTHOUR METER
M1
M4
1
2
3
4
MAIN FIELD AND INTERPOLE FIELD RELAY EQUIPMENT
CMP1
CMP2
CMP3
CMP4
IP1
IP2
IP3
IP4
F15X
F14X
F14R
F44R
F44X
F45X
AUTO-TRANSFORMER
F12
F13
F11
F32
F33
F31
INDUCTIVE SHUNT
EXC. FLD. 1
EXC. FLD. 2
EXC. FLD. 3
EXC. FLD. 4
WSB1
WST1
WSB2
WST2
M5
WSB3
WST3
POWER TIE
F24R
F24X
F25X
F35X
F34X
F34R
AMMETER TRANSFER CIRCUITS
A
M2
M3
CMP - COMPENSATING FIELD
IP - INTERPOLE FIELD
- REVERSER CONTACTS
WSB - WHEEL SLIP WARNING RELAY
WST - WHEEL SLIP TRIP RELAY

Figure 11. Schematic diagram of traction motor circuits, motoring operation

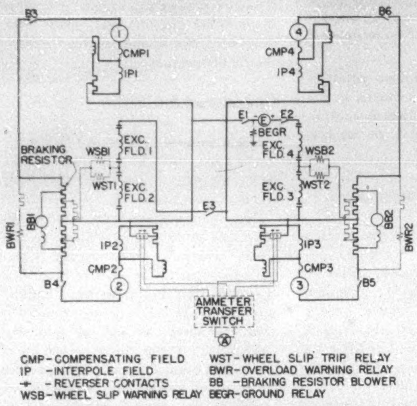
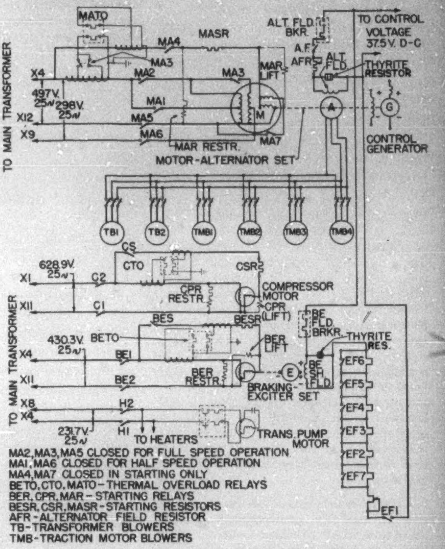

B3
B6
1
CMP1
IP1
4
CMP4
IP4
E1
E
E2
BEGR
EXC. FLD. 1
EXC. FLD. 4
BRAKING RESISTOR
WSB1
WSB2
WST1
WST2
EXC. FLD. 2
EXC. FLD. 3
E3
BWR1
BB1
BB2
BWR2
IP2
IP3
CMP2
CMP3
B4
2
3
B5
AMMETER TRANSFER SWITCH
A
CMP - COMPENSATING FIELD
IP - INTERPOLE FIELD
- REVERSER CONTACTS
WSB - WHEEL SLIP WARNING RELAY
WST - WHEEL SLIP TRIP RELAY
BWR - OVERLOAD WARNING RELAY
BB - BRAKING RESISTOR BLOWER
BEGR - GROUND RELAY

Figure 12 (above). Schematic diagram of traction motor circuits, rheostatic braking connections

Figure 13 (right). Schematic diagram of auxiliary circuits


MATO
MA4
MASR
ALT. FLD. BKR.
TO CONTROL VOLTAGE 37.5 V. D-C
A.F.
AFR
ALT. FLD.
THYRITE RESISTOR
MA3
X4
MA2
MA3
MAR LIFT
497 V. 25~
298 V. 25~
MA1
M
A
G
X12
MA5
X9
MA6
MAR RESTR.
MA7
CONTROL GENERATOR
MOTOR-ALTERNATOR SET
TO MAIN TRANSFORMER
TB1
TB2
TMB1
TMB2
TMB3
TMB4
CS
CTO
CSR
628.9 V. 25~
X1
C2
CPR RESTR.
COMPRESSOR MOTOR
CPR (LIFT)
X11
C1
BES
BESR
BE FLD. BRKR.
THYRITE RES.
430.3 V. 25~
BETO
BER LIFT
X4
BE1
EF6
BER RESTR.
E
BE SH. FLD.
X11
BE2
EF5
BRAKING-EXCITER SET
TO MAIN TRANSFORMER
X8
H2
EF4
X4
231.7 V. 25~
H1
TO HEATERS
TRANS. PUMP MOTOR
EF3
EF2
EF7
EF1
MA2, MA3, MA5 CLOSED FOR FULL SPEED OPERATION
MA1, MA6 CLOSED FOR HALF SPEED OPERATION
MA4, MA7 CLOSED IN STARTING ONLY
BETO, CTO, MATO - THERMAL OVERLOAD RELAYS
BER, CPR, MAR - STARTING RELAYS
BESR, CSR, MASR - STARTING RESISTORS
AFR - ALTERNATOR FIELD RESISTOR
TB - TRANSFORMER BLOWERS
TMB - TRACTION MOTOR BLOWERS

sistor is connected in series with a quarter-phase starting winding until the set is nearly up to speed. A balanced-beam type relay then cuts out the starting winding. This relay has a pick-up coil energized by the voltage across the starting winding, and a restraining coil connected to the motor line voltage. Thus it compensates for line voltage changes and allows starting over a wide range of applied voltage.

Low-Voltage Control Circuits

The master controller has a main handle to control acceleration, a reverse handle, a braking handle to control rheostatic braking, and a selector handle to control the motor-alternator set. Suitable mechanical interlocking prevents improper operation of these handles.

Control of the transformer tap contactors for acceleration is obtained through intermediary sequence relays, one for each contactor. Previous locomotives using the 3-preventive-coil system had contactors interlocked to prevent connecting more than one tap at a time to the same bus. The use of sequence relays simplifies the interlocking by reducing the number of circuits on each contactor. It also makes the interlocking more complete by precluding overvoltage on the preventive coils in case a contactor sticks closed. Since the relays operate on a small amount of power, they reduce the current carried by the train-line jumpers between locomotive units.

Protection

Circuit and apparatus protection for these locomotives centers in a transformer protective relay for the main transformer, and a ground relay for the traction-motor and auxiliary circuits. These are supplemented by wheel-slip relays for the traction motors and thermal overload devices for the auxiliaries.

Ground Relay

A ground relay of the motor-element type is connected, through an autotransformer, to taps on the main transformer secondary. Should a ground occur on the main transformer secondary; on the bus bars, tap contactors, or preventive coils; or any other equipment in the main circuit up to the traction-motor line contactors and auxiliary contactors, an auxiliary relay automatically inserts resistance in series with the ground-relay operating element, thus desensitizing it.

The ground relay has a set of *A* contacts which trip when the relay element rotates partway and a set of *B* contacts which operate when the relay travel is complete. In case of a ground on a traction-motor or 25-cycle auxiliary, the *A* contacts open and latch in the tripped position, de-energizing the traction-motor line contactors and auxiliary contactors. With these open, the grounded equipment is cleared and the relay returns to normal position. The *A* contacts may then be reset remotely from the operator's position. After the faulty circuit has been isolated, operation, possibly at reduced capacity may be continued.

Should the ground be on the transformer side of the various line contactors, as mentioned above, the ground relay will not stop after tripping its *A* contacts, but will continue around and close the *B* contacts, causing the auxiliary relay to operate. This connects the desensitizing resistor in series with the ground relay operating coil, causing the latter to return to its original position so that the *A* contacts may be remotely reset. This arrangement permits operation of the locomotive in an emergency with a ground between the main transformer secondary and the line contactors of connected apparatus, with the ground relay having been automatically desensitized.

Transformer Protective Relay

The transformer protective relay (*TPR*, Figure 11) combines primary and secondary overload and short circuit protection in a single device. It has both magnetic trip coils and a thermal element trip. The latter, energized from a current transformer in the high-voltage primary bushing, insures against the relay tripping on transient high currents, but permits it to operate when high currents persist.

Each trip coil is connected to a current transformer mounted on a tap where it comes out of the main transformer terminal board. These current transformers and their associated tripping coils are arranged to give tap to tap protection by instantaneous tripping on all currents in excess of the normal surges encountered when changing taps.

The trip coils, together with the thermal element, operate upon a common latch bar. When this latch bar is released by the operation of any of the relay elements, the magnet valve of the ground switch is energized, thus grounding the pantograph. Flow of current through the ground switch partially trips the ground switch relay (*GSR*, Figure 11) and isolates the pantograph control so that the pantograph cannot be lowered remotely. When the trolley circuit breaker opens, it cuts power off the trolley circuit to the locomotive and de-energizes the ground switch relay. This, then, completes its cycle and operates the down magnet valve, lowering the pantograph.

The transformer protective relay and ground switch relay are manually reset.

Wheel-Slip Protection

There are two systems of wheel-slip protection for the traction motors. One consists of a power tie connecting the mid-voltage points of both 2-motor series circuits. Note that for each circuit there are two motors in series and both are on the same truck, see Figure 11. Two current-element, wheel-slip relays are located in this power tie. Under heavy accelerating conditions there will be an appreciable weight transfer from the leading to the trailing axle in each truck. Hence, the leading axle will tend to slip first. For instance, if Numbers 1 and 3 axles are leading, the Number 1 motor slips, the voltage on Number 1 motor and, because of the power tie, Number 4 motor will rise. Also there is a corresponding decrease in the voltage on Numbers 2 and 3 motors. Number 1 motor current decreases because of slipping; Number 4 motor current increases; and Numbers 2 and 3 motor currents decrease. Part of Number 4 motor current then flows through the power tie to Number 2 motor. The results thus accomplished are:

1. Voltage build-up and tendency of the slipping axle to overspeed are minimized, making it easier to recover from a slip.
2. One of the more heavily loaded axles gets the current build up, and the more lightly loaded axles get a reduction.
3. The current reduction is minimized in the motor which is paired with the slipping motor.
4. Tractive effort is maintained at a high value, even though an axle slips.

If sufficient current flows through the power tie, the wheel-slip warn-warning relay, *WSB3*, operates a signal light and buzzer at the operator's position. If the slip is severe enough, the wheel-slip trip relay, *WST3*, will operate to open the motor line contactors and remove power from the traction motors. The master controller must be shut off to reset this relay. The second wheel-slip protective system is provided by a-c and d-c voltage relays, two in each 2-motor circuit. They operate on a difference in potential between the two motors in each group. One relay in each group is a warning relay, the other a trip relay. The contacts perform the same functions as described above for current relays. The voltage relays are particularly useful in rheostatic braking, or when one pair of motors is cut out as, in either case, the power tie circuit is open.

These systems also provide protection against overspeeding as a result of a slipped pinion on the traction motor.

Protection against overcurrent in rheostatic braking is obtained by a voltage relay connected across each braking resistor. When the permissible current is exceeded, these operate a warning light and buzzer at the operator's position.

Ground protection during rheostatic braking is furnished by a ground relay, *BEGR*, connected in the exciter armature circuit. When it operates, warning lights indicate in the operating cab and the braking power and exciter field circuits are opened. This relay is reset manually.

Thermal overload relays furnish protection for the 25-cycle a-c motors of the motor-alternator set, the braking exciter set and the air compressor. These trip out the motor line contactors on overload, and are manually reset. Trip-free manually-operated, thermal overload circuit breakers are used for overload protection of the heaters, Pyranol pump motor, alternator field, exciter field, and various control and light circuits. A magnetic-type series overload relay furnishes protection for the 37.5-volt control generator.

Protection against loss of traction-motor ventilation is furnished by two air-pressure relays, one for each two blowers. If one blower is shut down, the relay lights indicating lights in the operating cab.

Indicating lights in two groups are arranged to give local indication for a particular locomotive unit and over-all indication for all units operating in multiple.

Summary

These locomotives have been designed to meet certain basic requirements:

1. A unit size to give the most efficient use over a wide range of traffic conditions.
2. A type of electric equipment well proved on the Railroad and most suitable for its requirements.
3. High horsepower. The locomotive rates continuously at 625 horsepower per axle, 26.5 miles per hour. It will, however, accelerate trains up to 33 miles per hour at 25 per cent adhesion, corresponding to 1,300 horsepower per axle, and haul tonnage trains over ruling grades at 41.5 miles per hour at 16 per cent adhesion, 1,060 horsepower per axle.
4. Arrangement of equipment to give ready access for inspection and maintenance.
5. Higher availability and lower operating expense.
6. Lower first cost.

These requirements were met by modern design, improved manufacturing methods, the use of better materials, and the selection of parts which lend themselves to use on many other locomotive types. Thus the advantages of advanced design were combined with the benefits of repetitive manufacture. Evidence from operation to date indicates that these objectives have been attained.

References

1. The Single-Phase Commutator-Type Traction Motor, **Felix Konn.** *General Electric Review* (Schenectady, N. Y.), volume 35, April, May, and June 1932, pages 206–14, 275–79, 397–402.

2. Twenty-Five Years Progress in the Design of Traction Motors, **M. J. Baldwin.** *AIEE Transactions*, volume 68, part I, 1949, pages 132–37.

Discussion

W. A. Brecht (Westinghouse Electric Corporation, East Pittsburgh, Pa.): It would be interesting to know why so many auxiliary rotating machines have been used on this locomotive, as compared with the Westinghouse ignitron locomotives in comparable service on the Pennsylvania Railroad. I note that each General Electric unit has 12 motors or motor-generator sets, comprising a total of 15 machines, whereas each ignitron cab unit has only four motors and one motor-generator set for a total of six machines. The total weight of these six machines is less than the reported weight of the General Electric motor-alternator set which supplies power to the auxiliaries.

The authors make the statement in their paper that, "It has long been recognized that much more performance can be obtained from a straight electric locomotive traction motor than from a diesel-electric or motor-generator locomotive motor." This is only true in the sense that power transmission over a trolley is virtually unlimited. Actually, more performance can be built into a d-c traction motor than an a-c motor of equal dimensions. The standard Westinghouse diesel traction motor, which is used on the Pennsylvania Railroad ignitron locomotive rates 500 rail horsepower continuously as compared with 625 for the a-c motor described in this paper. However, it is commonly mounted on 42-inch wheels, as compared with 48-inch wheels required for the a-c motor, and is ventilated with only 2,700 cubic feet per minute of air as compared with 4,200. And the d-c motor delivers its full horsepower rating down to about half the locomotive speed that the a-c motor does.

With the dimensions and ventilation of the a-c motor, a d-c motor rating substantially more than 625 horsepower can be designed. And the delivery of full horsepower down to lower speeds means added performance, particularly in freight applications.

D. R. MacLeod (General Electric Company, Erie, Pa,): The paper presented on the a-c series type locomotive records another step forward in the development of a-c electrification in this country. Since the conclusion of World War II General Electric has built the heaviest electric locomotive in the world for the Virginian Railroad, and the longest single cab electric locomotive in the world for the Great Northern Railroad. Neither of these locomotives fits the current operating patterns on the New York, New Haven and Hartford or Pennsylvania Railroads electrifications nor is the Great Northern locomotive particularly suitable to the Virginian. We have been doing this sort of thing for half a century—custom-building electric locomotives to meet each customer's peculiar requirements. Usually, even the components were so different that it was impossible to achieve any economy in manufacture to say nothing of the tremendous outlay of engineering talent that each order required. We became convinced that some measure of standardization of electric locomotives was essential if electrification was to be extended in this country.

Early in 1949 the Pennsylvania Railroad ordered two locomotives from the General Electric Company and two from the Westinghouse Electric Corporation[1] and gave each manufacturer wide freedom of selection of design. The Pennsylvania Railroad exercised decisive control only in such matters as maximum weight that was to be carried on each driving axle, clearance requirements, and safety appliances. Each builder was to design the locomotive that its engineers believed to be the best freight locomotive for the Pennsylvania Railroad's electrified divisions. This forward looking attitude gave an opportunity to work toward a design that would be universally applicable to american railroads. It is interesting that different solutions for the same transportation problem have been offered by the two companies.

One important decision that had to be made was whether a 4-axle unit, a 6-axle unit, or an 8-axle unit should be used. This was a matter of weighing slightly lower cost per pound of adhesive weight for the units with the larger number of axles against flexibility of operation in Pennsylvania Railroad freight service and adaptability to other electrifications. We were also influenced by the knowledge that if a 4-axle unit could be built with a given type of electric equipment for 60,000 pounds on each driving axle, it would be relatively easy to meet this weight limitation on a 6-axle unit. We, therefore, decided on a 4-axle unit similar to the road-freight diesel-electric that had proved so adaptable to every railroad in the country. Westinghouse elected to build a 6-axle unit. General Electric placed the emphasis on simplicity of arrangement and on a minimum number of devices to maintain, as an aid to securing low maintenance costs. Westinghouse apparently decided that low maintenance could be best secured by using d-c traction motors. The writer believes that they were also influenced in favor of this step by the attractive prospect of building an a-c locomotive from mass produced diesel-electric parts. General Electric believed that the movement of freight at high speed over ruling grades was the principal factor in making possible the phenomenal performance of the Pennsylvania Railroad electrification during World War II. This experience, coupled with the advent of the powerful gas turbine electric locomotive convinced our engineers that the only specification for the electric locomotive under the demanding conditions of the Pennsylvania Railroad was to give it the ability to deliver high tractive effort at speeds several times as high as the diesel electric, and with a comfortable margin of superiority over the gas turbine electric.

REFERENCE

1. THE PENNSYLVANIA RAILROAD RECTIFIER LOCOMOTIVE, C. C. **Whittaker, W. M. Hutchinson.** *AIEE Transactions*, volume 71, part II, 1952, pages 37–49.

S. V. Smith (Pennsylvania Railroad, Philadelphia, Pa.): Four new 2-unit electric freight locomotives have been built for the Pennsylvania Railroad. The two new a-c locomotives described in this paper were placed in service in June 1951.

Maximum tonnage ratings for all the new locomotives have been established in accordance with each manufacturer's recommendations. When ready for service, the first locomotive of each type was assigned a program to haul revenue freight trains with tonnages as close as practicable to recommended maximum ratings between the principal freight terminals in electrified territory. Engineering representatives of the railroad and the manufacturer rode the locomotive throughout this program. Thereafter, the locomotive was pooled with 152 older electric locomotives and assigned in turn to haul freight trains with only such further special technical attention as the manufacturer saw fit to provide.

Maximum tonnage ratings issued for the new locomotives follow the pattern previously developed for older electric locomotives—momentum operation with helpers coupled behind on certain ruling grades. Ratings for the new a-c locomotives were calculated from design. With such ratings, no difficulty in handling assigned tonnages on ruling grades has been experienced thus far. Apparently the new a-c locomotives can handle more tonnage than the manufacturer has recommended.

All of the new electric freight locomotives have six modern features in common. They are: the *A*-type unit, simplified trucks with small driving wheels, permanent fire extinguishing systems, dynamic braking, all weight on drivers, and axle loading slightly in excess of 60,000 pounds per axle. The value of these features has been proven on modern diesel-electric freight locomotives. All types of electric locomotives previously built in quantity for the Pennsylvania Railroad were originally designed for and still operate at times in passenger service. The application of all these features could not be economically justified, even today, on new passenger electric locomotives.

Anyone who undertakes to compare different types of locomotives should bear in mind that the maximum tractive force which any locomotive can transmit is mechanically restricted by two related factors. The first is fixed and is the weight on drivers. The second, variable and highly controversial, is the adhesion which can be maintained between the surfaces of the driving wheels and the running rails. In quotations for new locomotives, high adhesion qualities are often quoted, but never guaranteed without restrictions.

The basic differences between the two new types of locomotives built for the Pennsylvania Railroad are related to size, type of electrical transmission system, and designed application of electrical transmission system. The rectifier locomotive, with six driving axles per unit, weighs about 54 per cent more than the a-c locomotive which has four driving axles per unit. With no other differences, the heavier locomotive would be expected to have proportionately higher tractive force and horsepower ratings.

Insofar as basic components are concerned, the electrical transmission system on the a-c locomotive follows the pattern of its predecessors. A pertinent characteristic of the transmission system is indicated by the absence of motor overload relays. With present tonnage ratings, the a-c traction motors appear to have sufficient capacity to transmit all the tractive force which the locomotive's adhesive limits will permit. The locomotive will slip before the motors become overloaded. No information on this performance with higher tonnage ratings has been presented.

The locomotive's adhesive limits seem to be effective from starting to 40 miles per hour on good dry rail and to 50 miles per hour on properly sanded wet rail. Traction motors appear to be inherently self-protecting at locomotive speeds above 40 miles per hour. The potential overload of a-c traction motors looks most attractive on paper. However, the statement that "full tonnage trains can be hauled at speeds considerably in excess of the continuous-rating speed when necessary to meet operating conditions" applies only to the extent that prevailing adhesive limits of the locomotive are not exceeded.

To date both types of locomotives have

indicated that they are capable of hauling freight more efficiently than the older electric locomotives or than a diesel-electric locomotive of comparable size. After proper tonnage ratings and operating restrictions are established, both types will be afforded ample opportunity to prove their conflicting claims of superiority as they work side by side in the same established electric freight service. The lower over-all cost per unit of transportation service performed will be convincing proof.

The Pennsylvania Railroad Ignitron-Rectifier Locomotive

C. C. WHITTAKER
MEMBER AIEE

W. M. HUTCHISON
MEMBER AIEE

Synopsis: Continued and intensive developments in perfecting the ignitron rectifier for broad application in the industrial and power fields revived interest recently in using it for railway work. Extensive tests in the laboratory and on an experimental railway car have demonstrated that a locomotive using ignitron rectifiers offered promise in combining the best features of the a-c and d-c electrification systems. The features of two such locomotives built for the Pennsylvania Railroad are described in the following paragraphs.

THE two 6,000-horsepower ignitron rectifier electric freight locomotives recently placed in service by the Pennsylvania Railroad mark a significant innovation in the development of railroad motive power.

It has long been recognized that the low-voltage d-c series-wound motor is well suited for traction purposes. It has advantages over a-c traction motors of relative simplicity in construction and of the ability to develop higher starting torques without exceeding limitations on commutation or heating.

On the other hand, the single-phase a-c system, as adopted by the Pennsylvania Railroad, results in economies in transmission and distribution systems, both in first and operating costs.

Development of Ignitron Motive Power

The development of the mercury-arc rectifier for converting alternating current into direct current offered an opportunity to combine the advantages of the two systems. As early as 1914, a mobile rectifier was applied to a 25-cycle 11,000-volt multiple-unit car operating on the New York, New Haven and Hartford Railroad. It was removed from service after 22,000 miles of operation because the rectifiers of that day had not been developed to the point where they were practical for such an application. They were of the multianode, pumped type, and with the seals then available difficulty was experienced in maintaining the vacuum. There were also frequent arc backs. With the advent of World War I, the idea was temporarily shelved.

In the 1920's, an alternate solution was developed in the form of the motor-generator locomotive. Such a locomotive converts high-voltage a-c power collected from the trolley to low-voltage direct current by a locomotive-mounted a-c transformer and synchronous motor d-c generator set. A number of such locomotives have been built and have proven to be successful in the low-speed, heavy-drag service for which they are particularly adapted. However, such locomotives are necessarily heavy and expensive.

It was the refinement of the ignitron type of rectifier in 1932, followed by its extensive application during World War II in the production of aluminum and magnesium, which turned attention again to the possibilities of a rectifier locomotive.

Extensive tests were made to develop the soundness of the application, first in the laboratory and later in the field on a trial multiple unit car. The problems and their solutions have been described in a previous paper.[1]

Those tests were so successful that in the spring of 1950 the Pennsylvania Railroad agreed to change an order for two a-c motored electric freight locomotives of a new type to ignitron rectifier-equipped locomotives.

These ignitron locomotives have several outstanding advantages:

1. They are applicable to the a-c trolley system with its high voltage power supply.
2. They provide the same flexibility of control as the conventional a-c locomotive, a control unequalled for a wide range of sustained horsepower output.
3. They secure the high sustained tractive effort of d-c traction motors.
4. They incorporate apparatus which is currently produced in large quantity.

Performance

Figure 1 illustrates the performance of the ignitron locomotive in terms of tractive effort and speed in comparison with a standard 3-unit, 4,500 horsepower, 3,700 rail horsepower, diesel-electric locomotive in widespread use today, and a 2-unit 5,625 horsepower, 25-cycle a-c locomotive. Each of these locomotives has 12 axles with a traction motor on every axle. The a-c motors are the most modern of the resistance lead type. The mechanical features of the a-c motor locomotives are identical with the ignitron locomotives.

These curves illustrate the fundamental

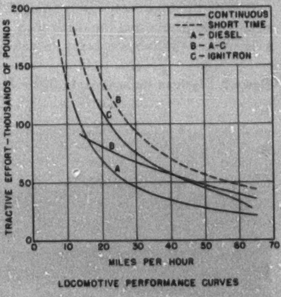


Figure 1. Locomotive performance curves

Paper 52-24, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21-25, 1952. Manuscript submitted October 23, 1951; made available for printing December 5, 1951.

C. C. WHITTAKER and W. M. HUTCHISON are both with Westinghouse Electric Corporation, East Pittsburgh, Pa.

Table I. Tonnage Ratings for a Typical Ruling Grade, 12-Axle Locomotives

	Trailing Tons	Mph
Rectifier	4,600	40
A-C Motor	4,700	
Diesel	2,700	
Rectifier	14,000	17
A-C Motor	9,000	
Diesel	8,300	

differences in performance of three locomotive types and point up the advantages of the ignitron rectifier type. It will be noted that the rectifier locomotive develops about 60 per cent more tractive effort than the diesel-electric locomotive over the entire speed range. The a-c locomotive has an advantage over both the rectifier and diesel-electric locomotive in the high speed range, as shown, but it will also be noted that in the low speed range it is the weakest. The rectifier locomotive produces continuously almost 50 per cent more tractive effort in the low-speed zone.

The advantage of the a-c locomotive in producing a short-time overload far in excess of its continuous rating is indicated. Actually, it also is possible to design for an equivalent overload with the rectifier system. However, there is no need for the added capacity since the continuous rating gives adequate performance in the low-speed range. The diesel-electric locomotive has a constant horsepower output, limited by the capacity of the diesel engine.

The information presented in these tractive effort curves can be translated into ability to pull trains. Table I shows tonnages that can be hauled by the three types of locomotives under comparison. These are based on a typical railroad division. The speeds shown occur on the ruling grade.

Dimensions and Ratings

One of these locomotives is equipped with 2-axle trucks and the other with 3-axle trucks, the purpose being to determine by service tests which type of truck will be the more suitable.

Figure 2 shows the 2(B-B-B) locomotive. The principal data and dimensions of the two locomotives having running gears 2(B-B-B) and 2(C-C) are shown in Table II.

Figure 3 shows the cab layout of the principal pieces of apparatus. The single aisle was dictated by the size of the transformer and the length of the ignitron cubicles and tap switch rack.

Ignitron Rectifier

In these locomotives the ignitron rectifier and its associated apparatus are made use of to convert energy received from an 11,000-volt 25-cycle a-c trolley wire to direct-current for the operation of conventional d-c traction motors.

The ignitron rectifier is a mercury arc device which employs an ignitor to initiate the arc at each positive half-cycle, and to permit the arc to cease at each negative half-cycle.

This article will not go into more detail relative to ignitron theory or structure because many papers have been written on this subject and are available as references.

Transformer

The transformer is inerteen-insulated and cooled by means of an air-blown radiator. Its design is essentially the same as that of a conventional a-c locomotive transformer except that the secondary is provided with a center tap with accelerating taps on both sides of the center. Because but one-half of the secondary is used for each half-cycle, the secondary winding is approximately 45 per cent larger than it would be for a single-phase series motor design. The inerteen is circulated by means of a pump built into the transformer tank. This transformer primary rates 4,200 kva at 11,000 volts. The tank is form fit to minimize the volume of inerteen required.

Figure 2. Ignitron locomotive, 11,000 volts, 25 cycle alternating current. 6,000 horsepower 2(B-B-B)

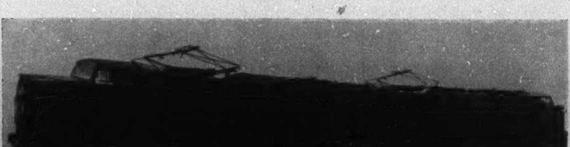


Table II. Principal Data

	Wheel Arrangement 2(B-B-B)	Wheel Arrangement 2(C-C)
Line voltage, a-c	11,000	11,000
Phase	single	single
Frequency	25	25
Number of driving axles	12	12
Weight—total, pounds	751,000	728,000
Weight on drivers, pounds	751,000	728,000
Weight per driving axle, pounds	62,500	60,666
Continuous rating at rail		
Tractive effort, pounds	132,000	132,000
Speed—mph	17	17
Adhesion—per cent	17.6	18.1
Starting tractive effort pound—25 per cent Adhesion	187,750	182,000
Maximum operating speed, mph	63	63
Number of traction motors	12	12
Driving wheel diameter, inches	44	44
Gear ratio	15:68	15:68
Air brake equipment	type 8-EL with D-24 brake valve	
Number of compressors	2	2
Compressor air displacement—total cfm	444	444
Type of control	Battery-electro-pneumatic	
Minimum curvature—coupled to train, degrees	16	16
Minimum curvature—locomotive only, degrees	21	18, 30′
Length over-all—between knuckles	124′ 0″	124′ 0″
Total wheel base	104′ 5″	104′ 5″
Rigid wheel base	9′ 6″	16′ 10″
Width over-all	10′ 1½″	10′ 1½″
Height over pantograph locked down	15′ 0″	15′ 0″
Height over cab	14′ 7⅜″	14′ 7⅜″
Width over cab	9′ 10″	9′ 10″
Track gauge	4′ 8½″	4′ 8½″
Coupler height	34″	34″
Clearance over rail—new wheels	5¼″	5¼″

Control Scheme, Motoring

As will be seen from the main schematic diagram, Figure 4, the transformer secondary is provided with a mid tap, d-c return, and a series of accelerating taps on each side of the mid tap. These taps feed anode busses through preventive coils. Each anode bus feeds an ignitron tube through an anode circuit breaker and anode reactor. These two tubes supply full-wave rectified direct current to a single traction motor. From each individual motor, the circuit continues through a d-c choke reactor, which limits the ripple to approximately 30 per cent, a motor switch, and a cutout switch back to the mid tap.

All six traction motors per A unit are similarly connected in parallel relation, in which connection slipping and spinning of drivers is minimized. In motoring there are 35 notches total. The sequence of switches is given with the main schematic diagram.

Phase delay of ignition to the tubes is made use of to increase the number of notches with a given number of transformer taps. This delay is varied up to 30 degrees in motoring and 50 degrees in braking. Figure 5 is a front view of the tap switch rack. This cubicle is entirely closed when in operation except for the arc box openings. It is ventilated by approximately 3,000 cubic feet per minute of air, which is expelled through the arc boxes. The contactors on the top of the box are starting and running contactors for the auxiliary motors and include a heater contactor.

Traction Motors

The traction motors are type *370-DZ*, each rated 500 horsepower. These motors have six main and commutating poles. Both armature and field windings are silicone insulated. Roller bearings are used on the armature shaft and sleeve bearings on the axles. These motors are the same as are used on Baldwin-Westinghouse and Fairbanks Morse Diesel road locomotives.

Auxiliaries and Ventilation

There are but three blower motors on each A unit. The traction motor blower ventilates the traction motors and supplies air for the windshield defrosters. It is a 1,450-rpm induction motor driving a Number $4^1/_2$ Buffalo Forge Niagara Conoidal runner requiring 43 horsepower.

Ventilation is by means of a duct in the underframe which feeds air into each 2-axle truck through the center pin and from the bolster to each motor.

On the locomotive using the 3-axle trucks, air is fed to each motor by means of a sliding high-hat arrangement on the underside of the duct directly to the motors.

The transformer blower is driven by a motor generator set. The motor of this set, which rates 29 horsepower, 1,450 rpm, also drives the control generator, a 45-volt 60-ampere machine and a centrifugal pump which circulates the cooling water for the ignitron tubes.

This blower ventilates the transformer radiator, the tap switch group, the d-c chokes, and the motor reactors.

The radiators in the cooling system for the ignitron tubes are ventilated by an induction blower motor rated 35 horsepower, 1,450 rpm. This motor drives an axiflow fan and draws air through the radiators and then passes the same air through the dynamic brake resistors mounted directly over the fan. The air is expelled through the roof. When the dynamic brake is used, shutters are opened by a pneumatic cylinder to bypass the radiators.

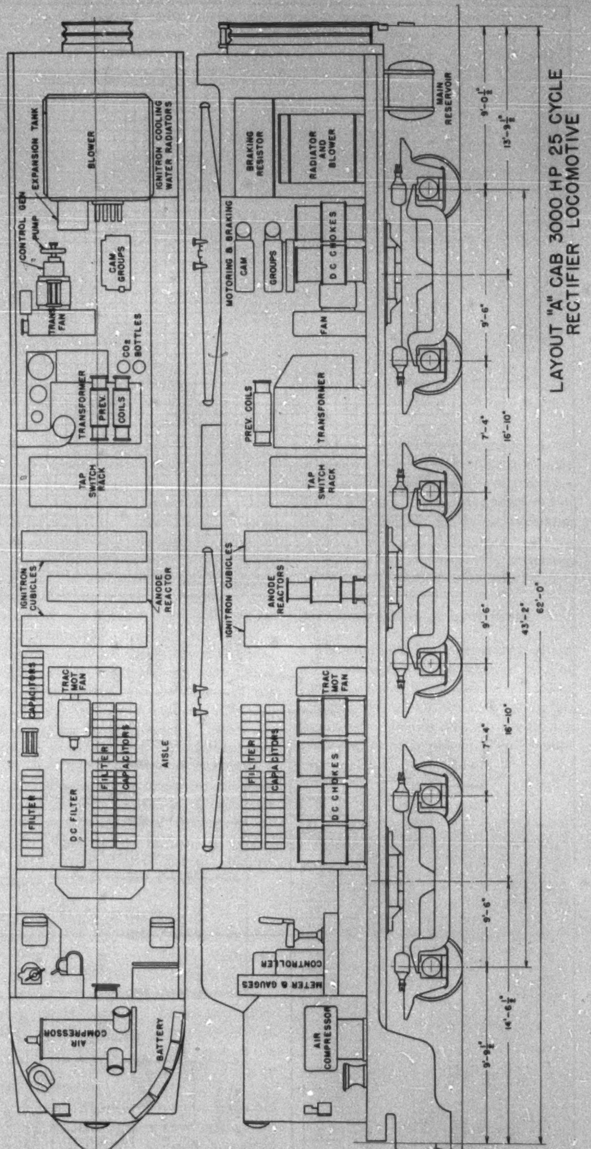

LAYOUT "A" CAB 3000 HP 25 CYCLE
RECTIFIER LOCOMOTIVE
EXPANSION TANK
BLOWER
IGNITRON COOLING WATER RADIATORS
CONTROL GEN
PUMP
CAM GROUPS
TRANS FAN
CO2 BOTTLES
TRANSFORMER
PREV.
COILS
TAP SWITCH RACK
IGNITRON CUBICLES
ANODE REACTOR
CAPACITORS
TRAC MOT FAN
FILTER
DC FILTER
FILTER
CAPACITORS
AISLE
AIR COMPRESSOR
BATTERY
BRAKING RESISTOR
RADIATOR AND BLOWER
MAIN RESERVOIR
MOTORING & BRAKING
CAM
GROUPS
DC CHOKES
FAN
PREV. COILS
TRANSFORMER
TAP SWITCH RACK
IGNITRON CUBICLES
ANODE REACTORS
TRAC MOT FAN
FILTER
CAPACITORS
DC CHOKES
CONTROLLER
METER & GAUGES
AIR COMPRESSOR
9'-0½"
13'-9½"
9'-6"
7'-4"
16'-10"
9'-6"
43'-2"
62'-0"
7'-4"
16'-10"
9'-6"
14'-6½"
9'-9½"

Figure 3. Layout A unit, 3,000 horsepower 25 cycle rectifier locomotive

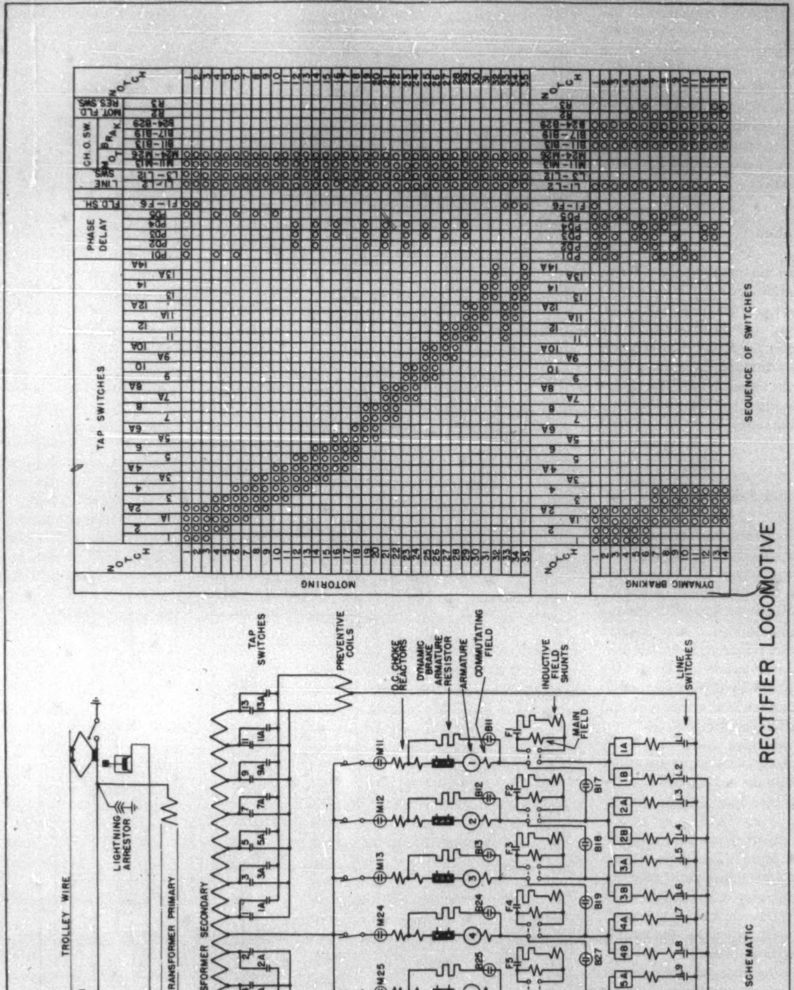

RECTIFIER LOCOMOTIVE
MAIN CIRCUIT SCHEMATIC
TROLLEY WIRE
PANTOGRAPH
PANTO GROUNDING SWITCH
PANTO LOCKDOWN & GROUND
LIGHTNING ARRESTOR
TRANSFORMER PRIMARY
GROUND
TRANSFORMER SECONDARY
TAP SWITCHES
PREVENTIVE COILS
D.C. CHOKE REACTORS
DYNAMIC BRAKE ARMATURE RESISTOR
ARMATURE
COMMUTATING FIELD
INDUCTIVE FIELD SHUNTS
MAIN FIELD
LINE SWITCHES
MOTOR CUTOUT SWITCHES
AMMETER SHUNT
MOTOR OVER-LOAD RELAYS
DYNAMIC BRAKE FIELD RESISTORS
IGNITRON TUBES
ANODE REACTORS
ANODE OVERLOAD RELAYS
SEQUENCE OF SWITCHES
MOTORING
DYNAMIC BRAKING
NOTCH
TAP SWITCHES
PHASE DELAY
FLD SH
LINE SWS
CH.O.SW.
MOT
BRAK
MOT. FLD RES SWS

Figure 4. Main schematic diagram and switch sequence of a rectifier locomotive

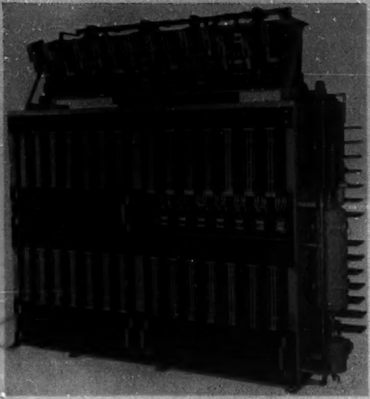


Figure 5 (left). Front view of tap switch rack

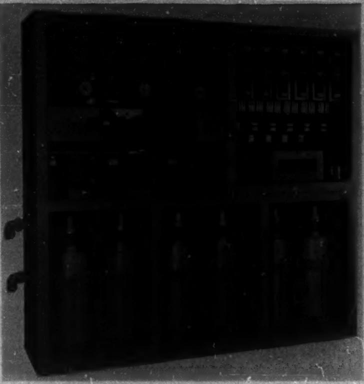


Figure 6 (right). Ignitron cubicle

Ignitron Cubicles

Figure 6 shows a view of an ignitron cubicle. There are two cubicles per locomotive unit. Each ignitron tube is mounted by a resilient structure which holds the tube at approximately its center of gravity. The mounting structure itself is fastened to insulating bolts, thus leaving the cathodes insulated from ground. These tubes are standard except for some special mercury baffling. At the upper left are seen three firing circuit controls, and at the upper right are six anode circuit breakers. Each circuit breaker must be capable of opening an arc back on an ignitron tube.

The most common trouble that has been associated with ignitron tubes in stationary service has been with arc backs. An arc back occurs when a cathode spot forms on the anode and a short circuit results. In a multiphase installation, the other phases feed power into this short circuit and the result may be a current flow of 50,000 amperes or more. It also is possible for a d-c load having stored energy in a shunt field to feed power back into the short circuit.

While using single phase for the railway application, no heavy arc backs have been experienced in several years of testing in our laboratories and two years' operation of the experimental multiple-unit car. If an arc back occurs, it is immediately extinguished at the next zero point on the voltage wave, and being a single-phase circuit, there is no other voltage present to reignite the arc independent of the ignitor itself. Since the load is composed of d-c series motors, there is no possibility of reverse power from this source.

Protective apparatus is provided to short the ignitor circuit and open the anode circuit breaker magnet valve circuit when an overload occurs. This results in the overload current being interrupted within the ignitron tube because the overload trip operates in one-half cycle, and the anode circuit breaker takes one and one-half cycles to open. Therefore, after an overload operation, the anode circuit breaker opens without an arc. This anode circuit breaker shown in Figure 7 with arc box removed is a new design rated 1,200 amperes alternating or direct current. It makes use of die cast parts instead of sand castings. It has air-core blowout coils. These have proved especially effective on direct current because of no residual flux and on either alternating or direct current have the feature of forcing the arc to the center of the arc box instead of burning arc box sides. The front of the coils is braced against magnetic forces and vibration by the arc chute guide.

Control Scheme, Braking

When the dynamic brake feature is used, the traction motor fields are excited by using two ignitron tubes to provide direct current. These two tubes then feed all motor fields in series through a stepped resistor and then through one of the d-c chokes back to the mid tap on the transformer.

Connected across each motor armature and commutating field is a braking resistor of fixed value. Fourteen notches in braking are obtained by the use of two taps on each side of the mid point of the transformer secondary, by varying a series resistor, and by utilizing phase delay on the two ignitron tubes.

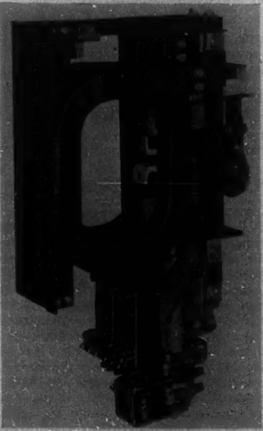
Figure 8 shows a single frame of the braking resistor. This resistor dissipates the energy supplied to it by one armature when braking. An over-temperature thermostat is provided in each braking circuit.

Figure 9 shows one of the motoring-braking cam switch changeover groups.

Air Compressors and Air Brake

The air brake is a combination of the *8-EL* type with a *DS-24* brake valve. The compressor is a Westinghouse Air Brake Company type *3-CD* driven by an induction motor. During the starting period, the compressor is unloaded and pumps

Figure 7. Anode circuit breaker, arc box removed



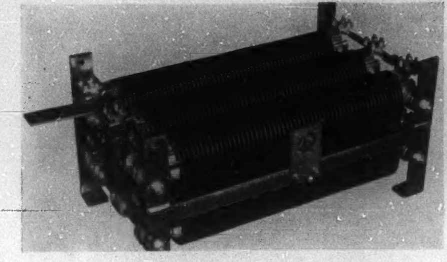


Figure 8. Dynamic brake resistor for one traction motor armature

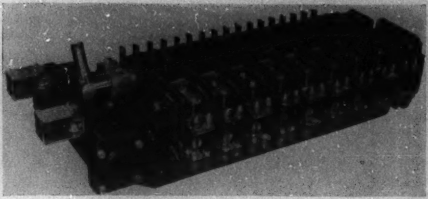


Figure 9. Cam switch change-over group for motoring and braking

against pressure only after it is up to speed. At 750 rpm, each compressor supplies 222 cubic feet per minute displacement.

Carbon Dioxide Protection

Carbon dioxide protection against fire is provided by four 50-pound tanks of CO_2. The apparatus compartment in each A unit is provided with a 6-foot fire wall adjacent to the aisle to retain the gas when expelled around the electrical equipment. Several discharge nozzles are located at strategic points. The blowers are shut down by the discharge of CO_2. While the transformer blower is slowing down, the reactors ventilated by this blower are filled with CO_2. The traction motor blower has its intake shut off completely when the CO_2 is discharged. Gas may be discharged at either side of the locomotive inside the cab or at either side of the locomotive from ground level.

If a fire is not extinguished by the gas in the unit where the fire occurs, the gas from the other unit may be used by opening a cock at the rear end of the unit not in trouble.

In addition to the preceding, there are portable CO_2 extinguishers.

At the time this article was written, the first of these two locomotives was undergoing tests preliminary to road acceptance tests by the Railroad. These tests indicated that the locomotive had ample margin in both motoring and braking performance. The value of the feature of having all motors in parallel was demonstrated on one test when, because of oil on the tread of the lightest-loaded axle, this axle continued to slip and balanced at a speed of 20 miles per hour while the locomotive was running at 16 miles per hour. Consequently, no slip relays have been provided.

In order to meet a satisfactory delivery date on these two ignitron locomotives and at the same time do some research work as to a-c and d-c filter requirements, it was thought best to build and equip the locomotives with the maximum size of filters that was thought might be required. Then, at a later date, after tests had been made in service, any excess filter capacity would be removed. This decision resulted in a somewhat crowded layout.

Our tests to date indicate that we can remove the d-c filter and approximately one-half of the a-c filter. This will result in more accessibility and will remove approximately 11,000 pounds of capacitors and reactor.

Conclusion

Because these locomotives are just going into service, no operating results are yet available. Continued operation is expected to demonstrate the advantages of this type of locomotive on existing single-phase a-c systems.

Its application is not limited, however, to 25-cycle operation, but it is equally adaptable to frequencies of 50 or 60 cycles. The principal deterrent to railroad electrification in this country is its high first cost. The ability of a rectifier locomotive to operate at commercial frequencies, as compared to a frequency of 25 cycles, necessitated by the limitations of a-c motor design, offers the possibility of reducing electrification costs and widening the field in which it is economically justified.

The two locomotives which have been placed in service by the Pennsylvania Railroad are therefore of interest, not only because of their advantages as a new type of motive power in this particular application, but also in the potentialities they have for the extension of electrification in the future.

Reference

1. Rectifier Type Motive Power for Railroad Electrifications, L. J. Hibbard, C. C. Whittaker, E. W. Ames. *AIEE Transactions*, volume 69, part I, 1950, pages 519–524.

Discussion

D. R. MacLeod (General Electric Company, Erie, Pa.): The papers by Mr. Gowans, Mr. Widell, and Mr. Bredenberg[1] on the a-c series type locomotive, and by Mr. Whittaker and Mr. Hutchison on the a-c rectifier type locomotive, record another step forward in the development of a-c electrification in this country. Since the conclusion of World War II General Electric has built the heaviest electric locomotive in the world for the Virginian Railway, and the longest single cab electric locomotive in the world for the Great Northern Railway. Neither of these locomotives fits the current operating patterns on the New Haven or Pennsylvania electrifications, nor is the Great Northern locomotive particularly suitable to the Virginian. We have been doing this sort of thing for half a century, custom-building electric locomotives to meet each customer's peculiar requirements. Usually even the components were so different that it was impossible to achieve any economy in manufacture to say nothing of the tremendous outlay of engineering talent that each order required. We became convinced that some measure of standardization of electric locomotives was essential if electrification was to be extended in this country.

Early in 1949 the Pennsylvania Railroad ordered two locomotives from the General Electric Company and two from the Westinghouse Electric Corporation and gave each manufacturer wide freedom of selection. The railroad exercised decisive control only on such matters as maximum weight that was to be carried on each driving axle, clearance requirements, and safety appliances. Each builder was to design the locomotive that its engineers believed to be the best freight locomotive for the Pennsylvania electrified divisions. This forward-looking attitude gave an opportunity to work toward a design that would be universally applicable to American railroads. It is interesting that different solutions for the same transportation problem have been offered by the two companies.

One important decision that had to be made was whether a 4-axle unit, a 6-axle unit or an 8-axle unit should be used. This was a matter of weighing slightly lower cost per pound of adhesive weight for the units with the larger number of axles against flexibility of operation in Pennsylvania Railroad freight service and adaptability to other electrifications. We were also influenced by the knowledge that if a 4-axle unit could be built with a given type of electric equipment for 60,000 pounds on each driving axle, it would be relatively easy to meet this weight limitation on a 6-axle unit. We, therefore, decided on a 4-axle unit similar to the road-freight diesel-electric that had proved so adaptable to every railroad in the country. Westinghouse elected to build a 6-axle unit. General Electric placed the emphasis on simplicity of arrangement and on a minimum number of devices to maintain, as an aid to securing low maintenance costs. Westinghouse apparently decided that low maintenance could best be secured by using d-c traction motors. The writer believes that they were also influenced in favor of this step by the attractive prospect of building an a-c locomotive from mass produced diesel-electric parts. General Electric believed that the movement of freight at high speed over ruling grades was the principal factor in making possible the phenomenal performance of the Pennsylvania Railroad electrification during World War II. This experience, coupled with the advent of the powerful gas turbine electric locomotive convinced our engineers that the only specification for the electric locomotive under the demanding conditions of the Pennsylvania Railroad was to give it the ability to deliver high tractive effort at speeds several times as high as the diesel-electric, and with a comfortable margin of superiority over the gas turbine electric.

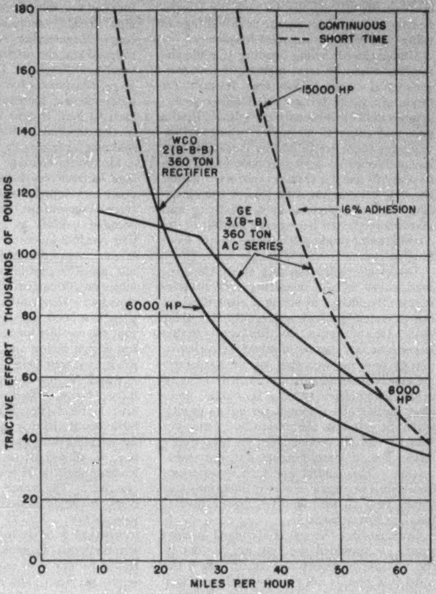


Figure 1. Comparison of speed. Tractive effort characteristics of 360-ton General Electric a-c series motor type and Westinghouse rectifier type locomotive

This paper and the paper by Gowans, Widell, and Bredenberg[1] are best compared by considering locomotives of the same weight on drivers. Figure 1 of this discussion compares a 360-ton a-c series motor type locomotive made up of three 4-axle cabs and a 360-ton ignitron type made up of two 6-axle cabs. The characteristics published in the two papers have been used. It will be noted that the comparison as shown on this curve is very different from Figure 1 of the paper. It should, therefore, be given very careful study. If the characteristics of the new 4,500-horsepower gas turbine electric locomotive were prorated to the same weight on drivers and the same number of axles as the locomotives that we are comparing, the speed tractive effort curve would be almost the same as that of the 6,000-horsepower 12-axle ignitron locomotive. In our opinion, this is too close for comfort for the low horsepower electric locomotive to compete for favor with railroad operating men.

Table I of the paper is apt to be misleading. The 8-axle locomotive described by the General Electric authors was designed to haul 10,000 adjusted tons at 30 miles per hour and actually hauled this tonnage at 40 miles per hour during the test runs over the same route on which the Westinghouse locomotives rates 14,400 tons. If the Westinghouse authors used the General Electric locomotives for their tabulation the comparison would be as follows:

	Trailing Tons	Mph
Rectifier	14,000	17
A-C Motor	15,000	17
Rectifier	4,600	40
A-C Motor	12,120	40

At the same time that the two locomotives for the Pennsylvania Railroad were built, General Electric built a third one of the same design to test on other a-c electrifications. This locomotive is now in operation helping to haul freight or passenger trains as required, over the 2.2 per cent grade on the Great Northern electrification through the Cascades. The purpose of this test is to determine whether the a-c series motor type can be used commercially and successfully in heavy duty mountain service. The motor generator type locomotive such as General Electric built for the Great Northern and Virginian electrifications since World War II, are ideally suited for these operations and there is no intention of proving that the a-c series motor type locomotive will suit their conditions any better. However, when the Great Northern and Virginian electrifications are extended, the factors that make the motor generator type so desirable will no longer be controlling and it will be of value to know that the types of locomotives that have proved ideal for the Pennsylvania operation, can also operate successfully over the heavy mountain grades found on the Great Northern and Virginian and that, therefore, this type can be applied to all four 25-cycle electrifications in this country.

The ignitron type locomotive is a very worth-while contribution to the art of electrification and we will watch its evolution on the Pennsylvania Railroad with sympathetic interest. It will be several years before a fair comparison can be made between these two types of electric locomotives which the Pennsylvania Railroad is pioneering.

REFERENCE

1. A NEW ELECTRIC LOCOMOTIVE FOR THE PENNSYLVANIA RAILROAD, F. D. Gowans, B. A. Widell, A. Bredenberg. *AIEE Transactions*, volume 71, part II, 1952, pages 27–37.

P. H. Hatch (Transportation Divisions, General Electric Company, Erie, Pa.): It is always somewhat difficult to judge a new design of locomotive which may be, as the authors state, over-refined in some respects at the start but which may later be capable of considerable simplification. The following remarks, therefore, are to be viewed in this light.

The salient feature of the new a-c locomotive is the use of d-c traction motors without the installation of rotating apparatus for converting alternating current to direct current. An inherent feature is that the alternating current may be at commercial frequencies as well as the conventional 25 cycles found on existing a-c electrifications in this country. Under certain conditions this might be of value for new electrifications.

It appears that consideration of the rectifier locomotive should be divided between: 1. its use on existing a-c electrifications where the power supply frequency is 25 cycles and; 2. its use on new a-c electrifications where there might be advantage in using power at commercial frequency.

Under the preceding heading 1, while the rectifier could be used instead of a motor-generator set on locomotives designed for mountain grade territory (though regeneration would be seriously complicated), the principal point of interest is what net advantages there are in adding a rectifier and its auxiliaries to a single-phase locomotive in order to use d-c traction motors. Either the over-all cost must be less or the operation, including performance, reliability and maintenance must be superior to that of the conventional single-phase series-motor locomotive.

Diesel electrification has given the railroads most of the operating and maintenance advantages of straight electrification but at considerably less over-all investment cost. One of the contributing factors to this lower cost is a degree of standardization in design and construction of diesel-electric locomotives hitherto unattained in any other locomotive types. It would seem that similar effort should, so far as practicable, be put into the design and construction of electric locomotives to take every advantage of simplification and standardization. This should apply to locomotives for either existing or new electrifications and would help to keep motive power investment as low as possible.

Granted that the use of standard or substantially standard d-c traction motors in the rectifier locomotive is an approach toward this goal is not the use of the rectifier and its associated equipment an offsetting factor?

Under heading 2, the rectifier locomotive appears to be a most interesting prospect for new electrifications where it is desired to use commercial-frequency alternating current, thereby eliminating the use of frequency changer sets as well as distribution apparatus built for noncommercial frequency. These advantages are not exclusive to the rectifier locomotive, but may also be secured with motor-generator set locomotives. There will still remain, however, the problem of using single-phase power from 3-phase supply.

In conclusion, it is believed that the development of the rectifier locomotive is of sufficient importance to warrant further appraisal of it in the light of these various questions. The Pennsylvania Railroad's service experience with the rectifier locomotive in comparison with existing electric locomotives should provide pertinent and valuable information for this purpose.

S. V. Smith (Pennsylvania Railroad, Philadelphia, Pa.): Four new 2-unit electric freight locomotives have been built for the Pennsylvania Railroad. The first ignitron-rectifier locomotive was placed in service early in November 1951. Delivery of the second ignitron-rectifier locomotive was made early in 1952.

Maximum tonnage ratings for all the new electric locomotives have been established in accordance with each manufacturer's recommendations. When ready for service, the first locomotive of each type was assigned a program to haul revenue freight trains with tonnages as close as practicable to recommended maximum ratings between the principal freight terminals in electrified territory. Engineering representatives of the railroad and the manufacturer rode the locomotive throughout this program. Thereafter, the locomotive was pooled with 152 older electric locomotives and assigned in turn to haul freight trains with only such further special technical attention as the manufacturer saw fit to provide.

Maximum tonnage ratings issued for the new locomotives follow the pattern previously developed for older electric locomotives—momentum operation with helpers coupled behind on certain ruling grades. The rectifier locomotive has been limited to 85 per cent of design rating. With this rating, no difficulty in handling assigned tonnages on ruling grades has been experienced thus far. When sufficient operating experience has been obtained, it seems probable that the rectifier locomotives will be assigned full design rating and operated on all ruling grades without help.

All of the new electric freight locomotives have six modern features in common. They are: 1. the *A*-type unit; 2. simplified trucks with small driving wheels; 3. permanent fire extinguishing systems; 4. dynamic braking; 5. all weight on drivers; and 6. axle-loading slightly in excess of 60,000 pounds per axle. The value of these features has been proven on modern diesel-electric freight locomotives. All types of electric locomotives previously built in quantity for the Pennsylvania Railroad were originally designed for and still operate at times in passenger service. The application of all these features could not be economically justified, even today, on new passenger electric locomotives.

Anyone who undertakes to compare different types of locomotives should bear in mind that the maximum tractive force which any locomotive can transmit is mechanically restricted by two related factors. The first is fixed and is the weight on drivers. The second, variable and highly controversial, is the adhesion which can be maintained between the surfaces of the driving wheels and the running rails. In quotations for new locomotives, high adhesion qualities are quoted, but never guaranteed without restrictions.

The basic differences between the two new types of locomotives built for the Pennsylvania Railroad are related to size, type of electrical transmission system, and designed application of electrical transmission system.

The rectifier locomotive, with six driving axles per unit, weighs about 54 per cent more than the a-c locomotive which has four driving axles per unit. With no other differences, the heavier locomotive would be expected to have proportionately higher tractive force and horsepower ratings. For purposes of technical accuracy and clarity, statements, tabulations, and curves relating to comparative performances and ratings should be confined to locomotives of the same size to which the same basic mechanical restrictions would apply. In this paper a rectifier locomotive and a diesel-electric locomotive of approximately the same weight appear to have been compared, inadvertently perhaps, with an a-c locomotive some 200,000 pounds lighter.

The total weight of each rectifier unit is about 7 tons less than that of the Pennsylvania's class *P5a* electric locomotive, which has 56 per cent of its total weight on drivers. Class *P5a* locomotives have for a number of years operated singly and in combinations of two or three in multiple to haul most of the freight trains in the Pennsylvania's electrified territory.

Insofar as basic components are concerned, the rectifier locomotive retains the transformer and tap switches used by its predecessors, adds ignitron tubes, and substitutes d-c traction motors already developed for diesel-electric locomotives.

A pertinent characteristic of the transmission system is indicated by the absence of wheel-slip relays.

Reduced tonnage ratings which now apply to the rectifier locomotive have afforded little opportunity for a practical demonstration of its adhesive qualities. Preliminary design curves and present motor operating restrictions indicate that some slipping may be experienced on sanded wet rail at speeds below 25 miles per hour. How effective the operation of six traction motors in parallel will be remains to be established. At higher speeds, maximum tractive forces which the d-c traction motors can transmit appear to be well below conservative estimates of the rectifier locomotive's adhesive limits. At speeds above 40 miles per hour, maximum tractive forces handled by the d-c motors should not be far below those obtained on the new a-c locomotives.

To date both types of locomotives have indicated that they are capable of hauling freight more efficiently than the older electric locomotives or than a diesel-electric locomotive of comparable size. After proper tonnage ratings and operating restrictions are established, both types will be afforded ample opportunity to prove their conflicting claims of superiority as they work side by side in the same established electric freight service. The lower over-all cost per unit of transportation service performed will be convincing proof.

P. A. McGee (New York, N.Y.): I wish to confine my comments to that part of Mr. Whittaker and Mr. Hutchison's paper which deals with a comparison between diesel and the ignitron-rectifier locomotive. In the first instance it is essential that there should be a proper understanding of the fundamental diesel locomotive characteristic of its full development of the pulling power of the d-c traction motors and the extensive application of same on the Pennsylvania Railroad.

The Pennsylvania Railroad has in operation or on order over 2,500,000 horsepower of diesel locomotives. This is the largest fleet of diesel locomotives in this country. I understand that when locomotives now on order are delivered, over 80 per cent of all traffic movement on the system will be by diesel-electric or single-phase straight electric locomotives.

The diesel locomotives employed are made up of units which permit a large range in tractive power per locomotive. By this means, the locomotive weight can be adjusted to suit variation in the profile of the different regions and variations in train size as dictated by traffic conditions over the entire system.

The total weight of the diesel freight units is on the driving wheels with a conservative

Table I. General Data of 3,000- or 4,000-Horsepower Diesel Locomotive with the 3,000-Horsepower Ignitron Locomotive

	Diesel Locomotive		Ignitron-Rectifier
Wheel arrangement	$B - B + B - B$	$B - B + B - B$	$B + B + B$
Horsepower rating	3,200 H-p	4,000 H-p	3,000 H-p
Horsepower at drive wheels, approximately	2,700 H-p	3,300 H-p	3,000 H-p
Total weight, maximum	500,000 Pounds	500,000 Pounds	375,000 Pounds
Starting tractive effort 25 per cent adhesion	125,000 Pounds	125,000 Pounds	93,750 Pounds
Continuous tractive effort	105,000 Pounds	105,000 Pounds	65,000 Pounds
Speed at continuous tractive effort	10 Mph	12 Mph	17 Mph
Maximum speed	65 Mph	65 Mph	63 Mph
Number of units	2	2	1
Number of motors	8	8	6
Number of compressors	2	2	1
Capacity of compressors	524 Cfm	524 Cfm	444 Cfm
Wheel diameter	42 Inches	42 Inches	44 Inches
Minimum radius curve, coupled to car	21 Degrees	21 Degrees	16 Degrees
Length over couplers	113 Feet	113 Feet	62 Feet

axle load which permits the operation of these units over all connecting lines. In an emergency, the Pennsylvania Railroad diesel locomotives, which may ultimately represent from 10 to 12 per cent of the total motive power on the United States railroads, could easily operate on rail lines between Boston and San Francisco.

The continuous tractive effort of many of the diesel freight units is in excess of 18 per cent of the total weight of the units. It is therefore natural that the Pennsylvania Railroad should be interested in the development of a single-phase straight electric freight locomotive which, within the limitations of the electric distribution system, would approximate the flexibility, weight efficiency, and performance characteristics and economy of the diesel units.

The Pennsylvania Railroad and the electrical manufacturers are to be congratulated on their efforts in this direction as exemplified in the interesting single-phase locomotives just delivered and described here today.

These locomotives are now under extensive tests on the Pennsylvania Railroad electrified lines. The results of these tests will be of great interest not only to a limited number of electrified railroads in the United States, but to a large number of foreign roads that seek our counsel, guidance, and financial support in the operation and development of their rail and transportation systems generally.

It is necessary, therefore, when introducing these experimental single-phase locomotives and recording their progress, to be conservative in their appraisal and honest in their claims.

We would like to be tolerant of the enthusiasm so necessary in an evangelist and proponent of something that is claimed to be new. However, I must take exception to the claimed performance comparison between diesel-electric and ignitron-rectifier locomotives presented by Mr. Whittaker and Mr. Hutchison.

Referring to a series of curves drawn to a very difficult scale, the authors state that a 6,000-horsepower ignitron locomotive, weighing 750,000 pounds with 12 d-c traction motors, develops 60 per cent more tractive effort than a 4,500-horsepower (3,700 horsepower at rail) diesel locomotive having the same weight, same traction motors, and geared for approximately the same maximum speed.

To complete the confusion caused by this very general statement, a table is presented referring to two unspecified grades in which they show that the ignitron locomotive can haul 4,600 and 14,000 tons on the respective grades against 2,700 and 8,300 tons respectively with the diesel locomotive.

I submit that there is absolutely no difference in the tractive effort or tonnage rating for any given value of adhesion between these locomotives if they have the same total weight and same continuous tractive effort. Actually, the diesel locomotive, although it is not indicated on the authors' obscure curve, has a 21 per cent greater continuous tractive effort than that shown for the ignitron locomotive. The distortion in the comparison submitted by the authors is, of course, easy to discern. Knowingly or unknowingly, they have confused torque, or ability to pull, with horsepower.

The standard rating cards for locomotives generally state the starting tractive effort and continuous tractive effort of a given locomotive with a tonnage rating over a particular division, founded on some accepted maximum adhesion value. Calculation and running tests indicate the horsepower required to fill a given schedule.

Locomotive designers know very well that tractive effort costs money. The standard diesel switcher, passenger, and freight units permit the purchase of tractive effort, or weight on drivers, if you please, at a given cost per pound. Manufacturers have more or less standardized on what they consider an economic balance between tractive effort and horsepower for switcher, passenger, and freight service. If, however, it is desired to increase horsepower without affecting the torque or tractive effort, it is a relatively simple and extremely cheap matter to add more diesel cylinders to your power plant and obtain any horsepower you want within a reasonable range with diesel locomotives.

There is no point in electrical manufacturers overemphasizing their ability to develop large horsepowers with low locomotive weight. When it comes to power development with low weight apparatus, high-speed diesel engines and gas turbines can give any reasonable answer required.

One of the largest purchasers of diesel electric transmission equipment from the authors' company has developed a line of road diesel locomotive units with the choice of three different diesel power plant sizes, namely 1,600, 2,000, and 2,400 horsepower per unit with approximately the same weight. These locomotive units employ the same traction motors and have the same continuous tractive effort rating with the same maximum speed with the three different power plants.

Had the authors considered this line of units, they would have found that their d-c traction motor, which they claim gives only a 60 per cent performance when supplied with diesel power, can, with a three unit (12 motor) diesel combination of 7,200 horsepower at the generator, give 6,000 horsepower at the rail. The 7,200-horsepower locomotive curve of necessity follows the ignitron curve included in the authors' paper. Here again this 7,200-horsepower diesel locomotive has the same tonnage rating on grades as its 4,800-horsepower companion locomotive, and it can deliver 21 per cent more continuous tractive effort than the ignitron locomotive.

The authors do not deal with the economics of the application. However, it is of interest to note that the increased cost of the 7,200-horsepower locomotive over the 4,800 horsepower is considerably less than 50 per cent.

So much for the diesel comparison presented by Mr. Whittaker and Mr. Hutchison. Now let us really compare the two types on a factual basis.

An analysis of the diesel units delivered to date to the United States railroads indicates that there is a very real use for 1,500, 1,600, and, to a lesser degree, 2,400-horsepower units in local freight, small passenger trains and branch line services. Two unit 3,000- or 3,200-horsepower locomotives can handle the majority of medium sized freight trains and large low-speed passenger and express train movements. Three-unit 4,500- or 4,800-horsepower locomotives can handle the majority of freight trains in this country and is the largest freight locomotive required on most railroads today. A four-unit 6,000-horsepower locomotive with approximately 500 tons on drivers represents probably the largest locomotive that can be economically employed with our present draft gear and brake equipment. Locomotives of this size are extensively employed on mountain divisions in the West and, to a lesser extent, on the shorter heavy grades in the East. In freight service, it is difficult to see just where you would employ a 6,000-horsepower locomotive with only 750,000 pounds on drivers. If such a locomotive was delivered with diesel power, I can see an immediate claim for more tractive power, either through more ballast or a trailing slug.

Referring now to the ignitron-rectifier locomotive made up of 3,000-horsepower 6-motor units: It is possible that the ignitron-rectifier does not lend itself to an economically smaller unit. If so, this is a decided limitation as universal use would certainly require a smaller unit for local freight, small passenger, and branch line work. The 3,000-horsepower unit with only six motors has not enough tractive power for the services now being performed by the 3,000-horsepower diesel locomotive. The step from a 3,000 to 6,000 horsepower, that is from 6 to 12 motors, is too large for a satisfactory range in locomotive sizes. In a final analysis, I feel that the 6,000-horsepower, 12-motor ignitron locomotive would, on American roads, generally just about compare with a 4,500- or 4,800-horsepower diesel locomotive. This conclusion is further supported by the requirements with dynamic or resistance braking. Regardless of horsepower, the operation on grade requires tractive effort for both hauling and braking. On the heavy grades in the West, the 6,000-horsepower locomotives weighing 500 tons are considered desirable for heavy dynamic braking services. In the East it is evident that a 3,000-horsepower locomotive in dynamic braking services can very profitably use 250 tons on drivers.

Another important advantage of the standard diesel units as developed, is the fact that for any given horsepower they carry more units with more compressor and air capacity. As an indication of the relative value of the two types of locomotives, Table I of the paper gives the general data of a 3,000- or 4,000-horsepower diesel locomotive with the 3,000-horsepower ignitron locomotive. This table very clearly brings out the aforementioned points mentioned.

H. S. Ogden (Locomotive and Car Equipment Department, General Electric Co., Erie, Pa.): Westinghouse is to be congratulated on its pioneering effort and accomplishment in the manufacture of a rectifier locomotive for heavy duty railway service.

When considering single-phase rectifier equipment of any kind, the question of the wave shape developed by the rectifier becomes important. The problem of filtering equipment has been approached by the authors and we can see from the contents of the paper that final answers are not yet available.

The application of the capacitors to the locomotive for filtering purposes also has a beneficial effect on the power factor of the locomotive. We would like to know what the locomotive power factor would be with the maximum capacitor equipment, and also what it would be with the minimum number of capacitors that it is contemplated will be necessary.

We note that the traction motors are Westinghouse 370*DZ*. We understand that the Westinghouse 370*F* motor is standard for use on Baldwin-Westinghouse and Fairbanks-Morse locomotives, and we wonder what significance there is in the difference in nomenclature.

The authors state that motor-generator type locomotives are necessarily heavy and expensive. The General Electric Company has built this type of locomotive for the Virginian and Great Northern Railways with approximately the same weight per axle as the Westinghouse rectifier locomotive and this weight included a substantial amount of ballast. Judging from the author's figures alone, no apparent weight advantage comes from the use of the rectifier, and we wonder if the authors would like to comment on this point.

The Section of the paper under, "Carbon Dioxide Protection" comments on the well-known fact that parallel-connected traction motors perform best under heavy tractive effort conditions when slipping is encountered. Here it is stated that a slipping wheel balanced at a speed of 20 miles per hour while the locomotive was running at 16 miles per hour.

It is not necessary to go to the elementary parallel connection with its attendant necessity of using high currents in the load supply. For instance, as was done on the locomotive described by Mr. Gowans, Mr. Widell and Mr. Bredenberg,[1] an equivalent parallel circuit was provided by connecting the midpoint of the two series circuits. Records of data obtained while original tonnage tests were being taken on the latter locomotive show a balancing speed of a slipping motor as 62 miles per hour with a road speed of the locomotive of 47 miles per hour. When expressed in per cent of the road speed, the slipping motor increase in speed checks quite closely with the percentage increase in speed observed by Mr. Whittaker and Mr. Hutchison.

We note that phase delay of ignition of the rectifier tubes is utilized to give intermediate steps between the main tap switch notches. Fundamentally, this kind of operation will affect the current ripple to the traction motors. Does the 30 per cent ripple mentioned apply to the rectifier action when there is no phase delay of the ignition, or when maximum phase delay of ignition is being utilized? We would expect that this type of control would require additional filtering capacity.

In closing, the writer would like to say that both of the locomotive developments that have been discussed here today will put those of us in the electric locomotive manufacturing business in a position to do a better job on existing electrifications, and also on further electrifications which will come when conditions are right.

REFERENCE

1. See reference 1 of D. R. MacLeod's discussion.

Gerald F. Smith (Ridgway, Pa.): This paper presents a new kind of electric locomotive and should be of interest not only in the United States, but abroad where there is considerable present interest in single-phase electrification at commercial frequency, principally 50 cycles in Europe.

The locomotive weight seems to be greater than was to be expected for a rectifier freight locomotive, in fact, it should not be difficult to build a motor-generator locomotive of equal capacity for the same weight.

A motor-generator locomotive would have a high power factor regeneration and low telephone interference. However, its efficiency would probably be lower. It would be of interest if further information on power factor, efficiency and telephone interference factor of the ignition locomotive were available.

It is a little difficult to understand why this locomotive was built in two units of six axles each, and with experimental running gear, instead of three simple units of four axles each. The increased flexibility of a 4-axle freight unit is one of the very valuable points of the diesel electric locomotive.

Looking at the performance curves, Figure 1 of the paper, curve *B* shows the continuous rating of an a-c locomotive which goes to a little over 90,000 pounds continuous tractive effort, or about $12^1/_2$ of the weight. Such a locomotive would be very useful if nothing better were available. However, the railroads have a right to expect, and can obtain, locomotives having continuous tractive effort of 18 or even 21 per cent of total weight. The dotted curve *B* shows the short time capacity of the a-c locomotive. However, no short time curve for the ignition locomotive is shown. This must have been an oversight as unquestionably, the d-c motor with nonfluctuating field is more capable under overloads than the a-c motor is.

Curve *A* shows the performance of a very widely used 12-axle freight locomotive. However, diesel locomotives are available which have higher horsepower (2,000 or 2,400 engine horsepower per cab) and can equal the performance of curve *C* of the ignition locomotive.

This then presents a question. Why are the railroads buying freight units of 1,500 or 1,600 horsepower when 1,200-, 1,600-, 2,000- or 2,400-horsepower units are available?

The hauling capacity of a modern locomotive (trailing tons which can be handled over the ruling grade) is determined by the weight on the drivers. The time to take the train over the road is dependent on the horsepower available. However, there is a point reached which determines how much horsepower can be economically used. Increased horsepower costs more, although not in direct proportion, as only the cost of power plant is increased. Higher capacity, however, cannot be fully utilized on the down grades and on return trips when the prevailing loads are substantially higher in one direction and hence, a point is reached where increased horsepower results in little time saving.

On this basis, the following suggestion is made to the Committee on Land Transportation:

Endeavor to have the Application Engineers of the railroads present papers on the application of locomotives to their particular railroads, indicating the economic phases of such applications with as many facts and figures as they feel free to give.

Considerable data have been presented in a general way in an International Communications Commission report.[1] Much more could be given by the individual railroads which could be of immense benefit.

I wish to emphasize another point in the paper; that is, the selection of parallel connected motors for these locomotives. The designers of this locomotive are to be congratulated on using this arrangement. The designers of diesel electric locomotives should waste no time in making this a feature of their locomotives, particularly on those of the higher power. Perhaps a paper on the relative operating performance of locomotives with parallel versus series connected motors could be obtained from the engineers of one of the Western Railroads.

REFERENCE

1. STUDY OF RAILROAD MOTIVE POWER. International Communications Commission (Washington, D. C.), 1950.

C. C. Whittaker and W. M. Hutchison: The wide variety of comments included in the discussions of this paper reveal a natural interest in this new form of electric motive power. It is not practicable in the limited space available here to answer in detail all of the questions which have been raised. However, the writers welcome the opportunity to comment briefly on several points.

The desirability of standardization of a-c motive power has been emphasized by Mr. MacLeod. Actually, one of the features of the ignitron locomotive is the degree of standardization which it offers. The motors are identical with those applied to diesel-electric locomotives in quantity. The ignitrons are the same as those also built in quantity for industrial applications. The adaptability of the rectifier locomotive to commercial frequencies offers the possibility of further standardization and simplification by elimination of frequency changers and by the utilization of distribution apparatus built for commercial frequencies. It is, of course, recognized, as indicated by Mr. Hatch, that there are still problems to be solved in obtaining conversion from three phase to single phase.

Some questions have been raised regarding the comparison in performance with other types of locomotives. As pointed out by Mr. S. V. Smith, the ignitron locomotive has many features in common with modern diesel-electric freight locomotives. For this reason, the diesel-electric freight locomotive in most general use, and with the same number of axles and about the same axle loading was selected as a bench-mark in evaluating the performance of this new form of electric motive power. The curves and tonnage table illustrate how, for a given speed, increased horsepower results in higher potential tractive force for hauling trains. The ratio of 6,000 to 3,700 horsepower is 162 per cent. Therefore, the potential tractive force at any specific speed is in the same ratio. How much of the potential tractive force can be utilized to pull trains is determined by the adhesion which can be maintained between wheel and rail. At the speeds shown in the tonnage table, that is, continuous rated speed of the ignitron locomotive, 17 miles per hour, and at higher speeds, including 40 miles per hour, the adhesion is within reasonable limits.

The ignitron locomotive, drawing its energy from the line, utilizes the full horsepower of the motors. The same motors, applied to a diesel-electric locomotive, can produce only as much horsepower as is provided by the prime mover. To provide 6,000 rail horsepower with a diesel-electric locomotive requires 7,200 engine horsepower.

The curve presented by Mr. MacLeod serves to illustrate that three of the 2,500-horsepower units, or a total of 7,500 horsepower, are necessary to match the tractive force of the 6,000-horsepower ignitron locomotive at approximately 21 miles per hour. While the increased horsepower would provide greater tractive force above this speed, because of the inherent characteristic of the a-c motor, it is deficient below that speed.

The tonnage values at 40 miles per hour, cited by Mr. MacLeod, appear to have been calculated on the basis of the emergency short time rating of the three 2,500-horsepower units and are, therefore, not comparable with the values for the ignitron locomotive which are based on the continuous rating.

As noted by Mr. S. V. Smith, operation of the ignitron locomotive was limited to 85 per cent of the design tonnage ratings for a trial period. These limitations have now been removed. The ability of the locomotive to effectively handle heavy freight trains over the electrified section of the Pennsylvania Railroad has already been well demonstrated. It is expected that as experience is accumulated the locomotive will further demonstrate its value.

Railway Electrification, Diesel-Electric Locomotives, and Some Future Aspects of Electric Traction

G. HULDSCHINER
MEMBER AIEE

Synopsis: Diesel-electric locomotives have replaced the steam locomotives in this country to a great extent and have been used in preference to railway electrification; whereas in Europe the tendency is still rather in favor of electrification. The following is a discussion for this diversity, which may be influenced by further technical developments.

RAILWAY electrification, born around 1900, had its heyday 30 years ago, here and abroad. In Europe, there are indications of its continued application; in this country, at least for the time being, the chances for further electrifications seem poor.

Railway electrification in this sense means, of course, the use of electric motive power produced in stationary power plants and fed into electric locomotives over some kind of contact system; as distinguished from the use of selfcontained vehicles like diesel-electric locomotives, in which the production and utilization of the electric power takes place in the same unit.

The diesel-electric locomotive has almost put a stop to further pure electrification in America; but no such marked tendency exists in Europe. The reasons for this divergency are partly technical and partly based on the differences in ownership, in financial structure of the railways, and in taxation practice.

One very clear reason is that oil, in this country, is abundant and cheap; whereas the countries practicing full electrification in Europe (Switzerland, Southern Germany, Austria, Italy, France, Scandinavia) have large resources of cheap water power, not much coal, and very little oil.

There is also a difference in the nature of the railway traffic. Densely populated areas, short distances apart, with corresponding heavy local movement and relatively light trains require a large number of locomotives per route mile; and as the diesel-electric locomotive is much more expensive than the relatively simple electric variety, at least in Europe, full electrification seems more tempting. On the other hand, for long railway lines across sparsely populated country, with little local traffic, the capital outlay for production, transmission, and conversion of current is preponderant; and the cost of a relatively small number of locomotives becomes an item of secondary importance.

Comparing the traffic features of both continents, it can be stated safely that most European railways belong to the first category, characterized by short runs, heavy local traffic, and limited train weight. On American lines, where long-distance through service with relatively little local movement and with heavy trains predominates and the cost of not too many expensive locomotives is more economical than the outlay for power plants, transmission, and conversion systems, the choice of diesels is a logical one. Where American railways had to meet dense traffic conditions similar to those prevailing in Europe, as on the New York, New Haven and Hartford line and on the New York—Washington D.C. line of the Pennsylvania system, full electrification was chosen.

Moreover, the financial structure of the railways, and consequently their ability to secure large new capital outlay, is quite different in America and in Europe. Electrification is costly business, and even if it pays in the long run—it does pay—money must be found for the initial investment.

Starting with the required electric power, a power plant is to be erected. In parts of the United States with large developed power production it is possible to buy power from the public utilities, a procedure favored in this country, but disliked by European state-owned railways. But this power must usually be converted into high voltage direct current or low frequency alternating current, and the cost of this conversion, whether centralized or allotted to many substations, is important. Long distance transmission lines and the trolley system are additional expense items. Straight electric locomotives are inherently cheaper than diesel-electric vehicles, but this difference is partly balanced when standardized diesels can be manufactured in large series. At any rate, the sum total of capital investment is usually much higher for full electrification, except for very dense traffic. There are also secondary expenditures peculiar to electrification only, such as the elimination of inductive interference and others, which shift the balance still further. The outlay of building capital imposes a large burden of interest, taxes, and amortization. Money must be found for the initial investment, very much money for full electrification, or a much smaller amount for diesel-electric locomotives; and the European governments, which own their railways, are in a better position to find the money than the private American railroad companies.

A state, even if not too prosperous, will find money for its railways; but a private corporation, mighty and powerful as it may be, but suffering from Government subsidized competition and already

---

Paper 52-25, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted June 29, 1951; made available for printing December 27, 1951.

G. Huldschiner is a lecturer in the Electrical Engineering Department, College of the City of New York, New York, N. Y.

heavily in debt, faces heavier interest charges and may have difficulty in procuring big sums.

There is also the matter of financing and taxation. In America the railways can secure new diesels by means of equipment trust certificates, but it is not yet possible to apply this type of financing to the investment of fixed property character (transmission lines, substations, contact systems) required for full electrification.

Moreover, in many states of the United States the railways are subjected to local taxes on their investment in fixed property, but not on their rolling stock. Such conditions do not exist in Europe.

The choice may also be influenced by another reason. The locomotive of a fully electrified railway is individually designed to meet the conditions of the line (longitudinal profile, speed, weight of trains, and density of traffic) with which the manufacturer has to comply. The electric locomotive is custom-built; perhaps to a smaller degree in America than in Europe, but still a vehicle selected and determined by the customer, as was the steam locomotive. Diesel-electric locomotives have become standardized for line production. Thus the two types of vehicles represent different concepts: the electric locomotive, built according to the specifications of the railway; and the diesel-electric locomotive, one of a series in which the characteristics have been determined by the maker, and which the railway has to adopt to its service. Line production is less developed in Europe and does not appeal so much to the European mind, especially not to the minds of state-owned railways with slightly autocratic tendencies which accept only their own standardization.

The heavy cost of electrification in America is partly due to the large weight of many locomotive types. In diesel-electric machines it can hardly be avoided because of the double conversion of energy. In purely electric locomotives the question as to whether or not they could be built lighter and cheaper seems to be justified, at least if the locomotives are destined for export and have to compete with European ones. A comparison between American and European locomotives shows a marked difference in weight. Many heavy American types of older design have a total weight of about 300 pounds per horsepower, some as much as 400. More modern types of powerful high-speed machines weigh 150–170 per horsepower continuous rating. The more recent locomotives of the Pennsylvania and the New York, New Haven and Hartford line weigh about 90 pounds per horsepower. European types are much lighter. One of the oldest machines of this kind, the Giovi line locomotives of the Italian State Railways, built about 45 years ago, weighed 64 metric tons with an hourly rating of 2,000 horsepower or 72 pounds per horsepower. The same or even smaller figures prevail for modern high speed, high power, single-phase locomotives in Germany and Switzerland. A very interesting extreme example is the latest type locomotive of the Loetschberg railway in Switzerland: single-phase $16^2/_3$ cycles, type B-B, maximum speed 80 miles per hour, equipped with 4 motors each 1,000 horsepower hourly rating, total weight 80 metric tons, corresponding to 45 pounds per horsepower. The most recent powerful electric locomotives for the Gotthard line in Switzerland show about the same relation.

It is true, that locomotive ratings and weights should not be compared without consideration of train weights, and American trains are much heavier than European ones. But on the other hand, many European lines have long ruling grades up to 3.5 per cent. By assuming that the horsepower output of locomotives per unit vehicle, is about the same in both continents, which seems true, and comparing the speeds, which are not much different, one must draw the conclusion that the performance in tractive effort should be about the same. The difference in weight might then be partly explained by the use of different coefficients of adhesion. American railways adopt in general 20 per cent adhesion, whereas in Europe 25 per cent seems normal and is sometimes exceeded.

The tendency to build light locomotives in Europe is of course also dictated by the European track structure, which limits the weight per driving axle to 20 metric tons. But if it can be done satisfactorily, there is no reason why it should not be done.

The very great weight of many American locomotives and the weight limitations per driving axle make a large number of drivers necessary. So the locomotive must sometimes be split into two or even three units. This results in duplication of parts, and as a consequence, in the almost outrageous weight of some of the older types of American locomotives.

It is true that a very heavy locomotive has, in general, better riding qualities than lighter vehicles. On the other hand, excess weight costs excess money in purchase and in nonrevenue ton-miles, and may also deprive the locomotive of an important safety factor. A correctly balanced locomotive should have a weight on its drivers equal to from four to five times the maximum tractive effort at the wheel-rims which the motors can develop without damage to the commutators. If the locomotive were heavier, the trailing load might be increased excessively with damage to the commutators.

European locomotives are not only lighter but the cost per unit weight is low, or at least it was before the war; this is mainly due to the lower cost of labor in Europe. Around 1930–1935 the Swiss and German manufacturers of electric locomotives could sell a main line machine for about 2.50 Swiss Francs per kilogram total weight. At the then prevailing rate of exchange, this corresponds to about 35 cents per pound. This, together with the difference in weight, indicates that the export of American locomotives may, in the future, be handicapped by strong European competition. Among the European countries, only Spain figures as a potential customer. But South America, South Africa, and India are markets open to competition with Europe.

As to pure electrification, the old controversy, direct versus alternating current will perhaps be revived in the light of possible new developments. In spite of all that has been written about this question, the controversy is still acute, especially as its aspect has changed several times and may change again. Up to about 1910, direct current and 3-phase current were predominant, with the exception of the New York, New Haven and Hartford single-phase electrification. Later on, single phase, developed by pioneers like Lamme in the United States and Eichberg and Behn-Eschenburg in Europe, opened the way to long line and heavy grade electrification in which the d-c system of moderate trolley voltage could hardly compete, handicapped as it was by the use of expensive conversion machinery of the revolving type. Around 1920 the question seemed rather stagnant, with a labile equilibrium between a-c and d-c partisans, but still with a slight accent in favor of alternating current. This balance shifted again with the development and perfection of the mercury arc rectifier, and the 3,000-volt d-c electrified railway presented so many advantages that today there is more parity between the systems. But improvements may change the aspect anew, and in this respect the possibilities seem more in favor of single phase.

Electrification with direct current on the contact system has reached a more or less saturated stage. The best way to reduce the cost, by the adoption of d-c trolley voltages substantially over 3,000 volts, does not appear feasible at the present stage of technique. There would not be any particular difficulty with rectifier equipped substations or with trolley lines. The problem would lie in the motor commutators, which can hardly stand more than 1,500 volts. A-c electrification is handicapped by the necessity of conversion from 3-phase industrial frequency to single phase at low frequency. Here decisive progress seems possible in two ways: equipment of locomotives with mercury arc rectifiers and d-c motors, or adaptation of the single-phase commutator motor to industrial frequency.

The first solution, the rectifier locomotive, combines all the advantages of high trolley voltage with the excellent qualities of the medium voltage d-c motor. All the locomotive control, except reversing, could probably be built in the low voltage side of the locomotive transformer; no different series-parallel combinations of the traction motors should be necessary and we would obtain a very simple control arrangement. On the other hand, the long and not very happy history of converter locomotives, dating back 50 years, forces adoption of a certain caution. Although the mercury arc rectifier is less bulky and heavy than the revolving machinery used in former types, it represents still a very unwelcome addition and its behavior on a locomotive is still an open question. Experimental locomotives of this type are tried in the United States and in Germany.

The other solution means adaptation of the single-phase commutator motor to 60 cycles (in Europe 50 cycles). Practical tests with 50-cycle motors were made in Germany before the war and the results were quite satisfactory. France, a country which adopted 1,500-volt direct current for its European lines 30 years ago (in the colonies—Morocco—they use 3,000-volt direct current) is now experimenting with 50-cycle single-phase locomotives on its secondary arteries in spite of the complication that these locomotives must also be capable to run on the d-c equipped main lines. This is no novelty, see New York, New Haven and Hartford line, but it certainly does not simplify matters and it proves that the French have enough confidence in the merits of industrial frequency a-c electrification to take complications into the bargain.

Of course, the adoption of industrial frequency alternating current creates other problems. Even if the necessity for frequency conversion were eliminated, the single-phase trolley line would have to be sectionalized and connected to different phases of the 3-phase high voltage system, which would cause some unbalance, or phase conversion from 3-phase to single-phase would be required. Both ways have been realized. Phase conversion has been accomplished by the use of mercury arc inverters and seems, if not simple, at least practically feasible. As to the sectionalization of the trolley line, the demand of the railway should constitute only a smaller part of the capacity of the 3-phase system; and if the density of service, the train weights, and the longitudinal profiles of the different sections are not too markedly different, the unbalances would not be serious. This requires very powerful 3-phase systems, and it has been successfully done on the New York, New Haven and Hartford. But it seems rather an exception, made possible by the proximity of the railway to a heavily industrialized region with its inherently large power production and consumption.

Another drawback would be the increased reactive drop in the trolley-return rail circuit. This would not be negligible and should not be overlooked. The total impedance drop at 60 cycles, including skin-effect in the rails, is more than twice as much as it is at 25 cycles. The per cent drop could be reduced by raising the trolley voltage, at present 11,000 volts in this country, and 15,000 volts in Europe. Increasing the trolley voltage to 22,000 or even 30,000 volts does not appear to present any serious difficulty; otherwise the distance between substations would have to be reduced.

Still another point to be considered concerns railway signalling. At present the track circuit used for signalling on 25-cycle electrified railways is fed with 60-cycle current, blocked at the end of each section by an impedance bond. This provides a barrier for the 60-cycle current, but allows the passing of the low-frequency rail-return current. The mere raising of the frequency of the traction current to 60 cycles would also result in the blocking of the traction current. In this case, the frequency of the signal current should be increased to 100 or 150 cycles with corresponding changes in the design of the impedance bonds and signal relays.

Although such changes would give rise to many new problems, the elimination of the frequency conversion represents so great an advantage that a thorough investigation is certainly justified. If technically good and not prohibitively expensive solutions can be found, the outlook for a-c single-phase electrification may be so greatly improved as to shift still further the choice between d-c and a-c electrification in favor of the latter.

A-c electrification is certainly very well suited for the heaviest kind of traffic, including dense local service, as the successful electrification of the Pennsylvania system proves; but it shows perhaps its best economical qualities for very long lines requiring only a relatively small number of locomotives per route-mile, because in this case the advantage of high trolley voltage is especially important. But lines of this character are also the favorite playground for diesel-electric locomotives; so progress in a-c electrification would also have its effect on the choice between electrification and dieselization.

It is not the purpose of this paper to make a general comparison between electrification and dieselization, and if advantages or disadvantages of one or the other system are mentioned, they refer to main line freight and passenger service and not to slow speed locomotives for large switching yards, for which kind of service diesel-electric locomotives are undoubtedly best suited.

No Discussion

A New Wheel-Slip Protective Scheme

R. M. SMITH
ASSOCIATE MEMBER AIEE

IMPROVED wheel-slip and wheel-lock protection for diesel-electric locomotives is a subject which has received considerable attention in the past few years. The system here described was designed to meet the proposed specifications of the Association of American Railroads. Its operation is based upon comparison of voltages which are proportional to the speed of the axles being protected. It will operate at the required 5-mile-per-hour speed differential between axles and will correct slipping in the specified manner. A thermally operated delay relay provides protection against abnormal conditions, such as a locked wheel. An additional feature is the provision of wheel-slip indication during dynamic braking. Oscillograms taken in actual train operation indicate that the system performs successfully.

Limitations of Present Systems

Equipment normally supplied on road locomotives depends upon a comparison of traction motor armature volts, field volts, or current as a primary detection of wheel slip. This arrangement inherently has a varying sensitivity, since the motor characteristics change with speed and the degree of saturation of the motor iron. This has the effect of leaving the locomotive with poor wheel-slip detection at high speeds where a combination of weak field and low motor current exists. If a motor, or entire unit in the case of multiple unit locomotives, is cut out, no protection is furnished against axle locking. Also, during dynamic braking wheel-slip indication is lacking.[1]

System Specifications

In a meeting of the Committee on Automotive Rolling Stock of the Association of American Railroads, a set of specifications for wheel-slip protection systems to cover these weaknesses was proposed.[2] These require that:

1. The relay panel shall consist of relays of the proper type that will operate with the axle equipment when the 5-mile-per-hour speed differential is reached between any axles under the unit.
2. The relay system is to be connected to the main generator excitation system so that, in the case of wheel slip, the excitation of the main generator and the engine speed will reduce, unloading the diesel engine until wheel slippage stops, and the engine will again slowly resume full speed.
3. The relay panel shall also have a timing device with a manual reset which will operate when slippage continues for approximately 30 seconds, indicating an abnormal condition, such as a locked wheel.

Description of System

A system meeting these specifications was designed and a number of experimental installations made. The objectives were, first, to collect fundamental data on the mechanics of wheel slip; and second, to prove the soundness of the apparatus. Each axle to be protected has a single-pole, double-throw switch mounted on the journal box, see Figure 1. This switch is actuated by a cam driven by the axle and so arranged that the switch contact goes through 5 cycles for each revolution of the axle. The switch terminals are connected to a control panel. The basic speed measuring circuit on the panel for one axle is shown in Figure 2. When the switch is closed in the position indicated, current flows through the switch contact, the current limiting resistor R_1, the capacitor C_2, and the load resistors R_2. When C_2 is charged, the current ceases to flow until the movable contact of the switch makes contact with the opposite pole. Capacitor C_2 then discharges through the short circuit thus formed, and capacitor C_1 is charged through the load resistors as already described. Capacitor C_3 acts as a filter to provide the equivalent of a d-c voltage across the load resistor.

The current through R_2, and hence the voltage across it, is proportional to the capacity of C_1 and C_2 and the frequency of switch operation. Hence the voltage across R_2 is a measure of the speed of the axle. If this voltage can be compared with similar voltages from other axles, wheel slip or slide can be detected. This comparison of voltages across the load resistors is accomplished by the circuit shown in Figure 3. Each load resistor is fed by a speed-measuring circuit, see

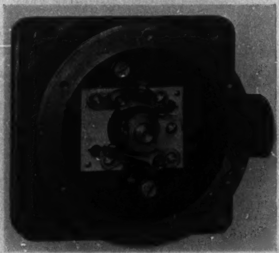
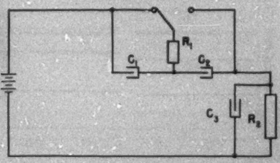


Figure 1 (left). Axle-driven switch unit arranged for mounting on journal box, with cover removed

Figure 2 (below). Basic speed-measuring circuit for one axle



Paper 52-35, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted October 23, 1951; made available for printing December 5, 1951.

R. M. SMITH is with General Electric Company, Erie, Pa.

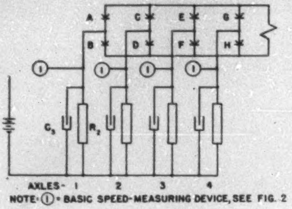


Figure 3. Circuit for comparing individual axle voltages

Figure 5. Oscillogram showing control of high-speed wheel slip by experimental equipment

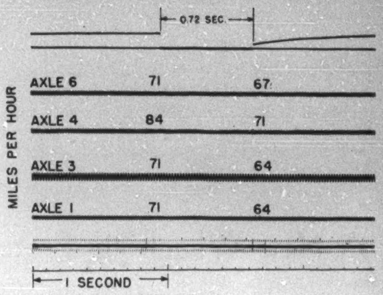


Figure 2, connected to its appropriate axle. The voltage from the load resistor of each axle is fed to the center of the rectifier combination. As long as the output voltages across all load resistors are equal, no current will flow. The output of these rectifiers is fed to a relay coil. When axle Number 1 slips the voltage across R_2 rises, causing current to flow to rectifier B. From there it flows through the relay coil, and then through rectifiers C, E, and G and the other three load resistors to the negative battery connection.

Actually this circuit measures the difference between the highest and lowest voltages in the system. In other words, it measures the speed differential between the fastest and slowest axles, whichever they may be, at any instant.

Locked-axle protection is obtained with this scheme under all conditions, including power on; coasting, power off; dynamic braking; and even with a motor cut out, or a unit cut out when in multiple operation.

Sensitivity

In practice it is necessary that the sensitivity of such equipment be tapered to provide greater sensitivity at low speed. This is necessary to allow for differences in wheel diameter and also to permit self-recovery to take place. The effect of wheel diameter is illustrated by the fact that at 90 miles per hour a 5-mile-per-hour differential in axle speed could also appear as a $2^1/_4$-inch difference in wheel diameter. This tapering in sensitivity is secured by adjusting the value of R_2 so that the volts per revolution at high speed is roughly one-half that at low speed. The ratio between high-speed and low-speed sensitivity can be varied by changing R_2, but a value of two to one appears to give good results without adjustment of the equipment for variation in wheel size.

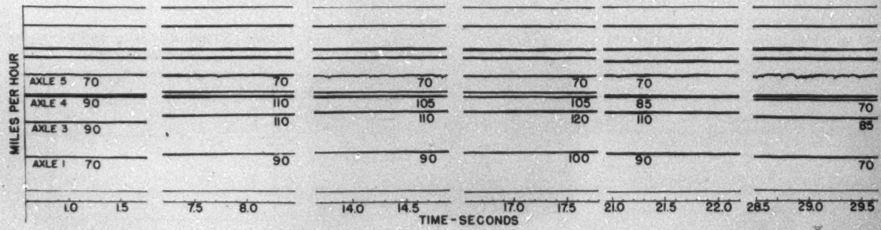
The case of a simultaneous wheel slip at speeds within 5 miles per hour of each other occurring on all axles, assuming no idler axles, would not be detected. Such a condition could be cared for, however, by train-lining wheel-slip equipment between units. This could be done with one wire connecting the negative side of the rectifiers on all units. The fact that simultaneous slipping of all motored axles does occur is shown by the oscillogram in Figure 4. Note, however, that although all axles are slipping, the individual axle speeds vary widely; being 100, 120, and 105 at a train speed of 70 miles per hour as read from the idler axle. The detecting equipment was indicating during this time, as shown by the top trace; but not connected so as to correct the slipping.

Functions

Section B of the Association of American Railroads Specifications[2] requires that the system correct wheel slip. On locomotives fitted with amplidyne excitation, this is accomplished by calling for minimum current limit. This is low enough to stop any normal wheel slip. When the slip drops below 3 miles per hour, full current limit is again called for, at whatever notch the engineman is operating. The capacitor-resistor combination across the limit-setting voltage, however, restores this higher setting gradually. This avoids causing a second wheel slip by too rapid a re-application of power. The time lapse for this delayed re-application is about 0.5 second. In practice, the operation is so rapid that, on normal starting slip or slip on heavy grades, the entire cycle from detection to restoration of full power is usually less than 2 seconds. A slip on very slippery rail, in which the energy storage is great in relation to the energy dissipation rate at the rail, will take longer.

The specification in Section C is met by a thermally operated time delay relay with an adjustable setting ranging from 15 to 30 seconds. The short duration of wheel slip on locomotives fitted with amplidyne control makes it possible to use a

Figure 4. Oscillogram showing sustained high-speed wheel slip on all axles



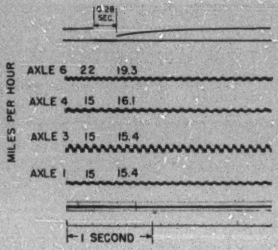


Figure 6. Oscillogram showing control of low-speed wheel slip by experimental equipment

15-second delay. On locomotives where it is necessary to call for a reduction of diesel-engine speed to idle, or where it is possible to obtain only a slow excitation recovery rate, there is a possibility of repeated slip building up a thermal storage in the time delay relay which would eventually cause a false trip. In this case, it may be advisable to increase the time delay.

Dynamic Braking Protection

When the locomotive is in dynamic braking, it does not seem advisable to control excitation by means of the wheel-slip apparatus. At such times the engineer should have full control of braking, and should simply be warned of wheel slip by the wheel slip light and buzzer rather than be deprived of his control over the braking effort. Wheel slip during dynamic braking seems to be rarer than during acceleration. On the other hand, it usually lasts longer when it does occur.

Since wheel slips in dynamic braking must be corrected by manual means, there is the possibility that repeated slipping will eventually trip the thermal relay. Such a condition may call for raising the time limit on this relay.

Operating Tests

The oscillograms taken on test runs and shown here illustrate wheel slip both with and without the experimental equipment controlling. The top trace shows the called-for current limit, and hence the tractive effort of the locomotive. The bottom dwell period is the time necessary for slip recovery. The exponentially rising line is the cushioned re-application of power. The trace for each axle has five peaks per revolution, corresponding to the operation of the cam-driven switch. A 60-cycle timing wave is used for reference.

The high-speed, simultaneous slip, see Figure 4, was taken without the equipment controlling. The entire sequence lasted about 40 seconds, during which time one or more axles were slipping. This slip occurred just at the beginning of a rainstorm. The rail was slippery enough to make it impossible to advance the throttle beyond the fourth notch without causing wild wheel slip. The engineman was aware of wheel slip from the indication of a separate light controlled by the experimental equipment, and the reading of the speedometer which was connected to a motored axle.

A second wheel slip at high speed with the experimental equipment controlling is shown in Figure 5. The detection occurred at 13-mile per hour differential in axle speeds, but the whole recovery cycle was fast because of the low motor torque available for starting wheel slip.

The action of the equipment at low speeds is shown in Figure 6. The rate of acceleration of the axle at the start is greater because of the higher torque available; but detection occurs at approximately 7. 1-mile-per-hour speed differential, so that recovery cycle is about the same. The control panel for this equipment is shown in Figure 7.

Figure 7. Control panel for wheel-slip protective equipment

Conclusions

The data taken have served to emphasize the high rates of wheel slip which can occur, and the extremely low factors of adhesion which must prevail to bring this about. Although it is difficult to determine speed closer than 1 or 2 miles per hour from the oscillograms, the fact that definite peaks occur at each contact closing removes the possibility of calibration error. This is especially helpful in field testing on locomotives, where conditions are not ideal for accurate calibration.

Experience with this equipment indicates that it is unusually fast in operation. As a result, it contributes to maximum locomotive performance by early detection of wheel slip. Its reliability in day-to-day service will only be demonstrated by longer operation.

References

1. LOCOMOTIVE WHEEL-SLIP AND WHEEL-LOCK PROTECTION, **R. M. Smith.** *AIEE Transactions,* volume 69, part II, 1950, pages 1154-57.

2. TWELFTH ANNUAL PROCEEDINGS OF LOCOMOTIVE MAINTENANCE OFFICERS' ASSOCIATION. Report of Committee on Automotive and Electric Rolling Stock—Subcommittee D, Washington, D. C.

No Discussion

Co-ordinated Transportation for Large Metropolitan Communities

E. E. KEARNS
ASSOCIATE MEMBER AIEE

TRAFFIC congestion in our large centers of population exacts a staggering toll in inconvenience and financial loss each year. Emphasis to alleviate this condition has taken the direction of expediting the movement of vehicles. Analysis shows the fundamental problem to be that of moving people. In the light of this, co-ordinated transportation systems, such as here described, offer most attractive possibilities.

Engineering Progress

If anyone were to select groups of men who have done most to improve our American way of life over, let us say, the past 30 years, engineers would certainly be nominated for a top place on the list. Menial tasks in home, field, and factory have been tremendously lightened and the time they consume greatly reduced. Radio, television, and other modern means of communication have greatly increased the volume and speed of the transmission of messages, music, and entertainment. Transportation of people and goods has made great strides in all forms; automotive, aviation and rail.

Traffic Congestion

Probably the greatest exception to this wonderful progress is found in the retrogression in speed and convenience of movement in and about our cities. We have spent many billions of dollars for wider streets, for limited access highways, for elaborate bridges, tunnels, overpasses and underpasses; for traffic controls and signs, and for facilities to park vehicles off the streets. In spite of these Herculean attempts to permit personalized transportation in the private automobile, traffic congestion in our cities grows steadily more acute.

The May 1950 issue of the *Reader's Digest*, under the caption "Ah, Prophecy," quoted the following from the 1899 *Scientific American:* "The improvement in city conditions by the general adoption of the motor car can hardly be overestimated—streets clean, dustless, and odorless—with light rubber-tired vehicles moving swiftly and noiselessly over their smooth expanse would eliminate a greater part of the nervousness, distraction, and strain of modern metropolitan life"; and from a recent issue of *Roads:* "Horse drawn carriages used to travel an average of 11.5 miles per hour in New York midtown traffic. Today the average speed of the automobile being used on the same streets is 6 miles an hour."

A friend of mine recently asked, "Shall we walk or have we got time to take a taxi?"

Space Limitations

The real essence of the problem is that we do not have, and never can have, enough streets, freeways, tunnels, bridges, and parking space to permit the movement of all people and goods over paved streets and highways. Take New York for example. It would require 12 to 15 6-lane, grade-separated parkways to provide the same passenger carrying capacity as a 4-track subway. Not only this, but where, oh, where would you park the cars delivered by such wide expanses of multiple freeways?

Recently I was shown a brand new building on the Miracle Mile of Wilshire Boulevard in Los Angeles. The far-seeing planners had heeded the admonition that it would take the floor area of another building of equal size to park only the cars of the workers in the building, if they all came in private autos. Of course, over and above those working in the building there are customers and clients who wish to come there too. Automobile parking space exceeding the total floor area of the building was provided beside and behind the building. The manager of the enterprise took me to the roof to see the result. The parking area was jammed and cars were cruising about the lanes to find parking space. His comment was, "We'd fill twice that parking area if we had it!"

Decentralization

Similar experiences are reported from all over the country. The larger the city the more acute the problem. In Detroit, the great Hudson department store finally found it necessary to set up branches on the periphery of the city. In their own words, "We had to concede the fight to the high and mighty automobile—after all Detroit spawned the millions of automobiles which have created the traffic congestion and parking problems plaguing all urban centers."

One very large corporation recently decided to move its office building out of downtown New York, out into the suburbs. In developing the project, parking space just double the floor area of the building is being provided, on the assumption that all their people will drive to the building. These folks have become realistic about the amount of parking area needed. Try to visualize what would happen if all New York shops and offices decided to follow this course. That sounds preposterous; but there is no mistaking the fact that this trend is very active at present. Primarily it results from our lack of a solution of the problem of

---

Paper 52-36, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21-25, 1952. Manuscript submitted October 23, 1951; made available for printing December 4, 1951.

E. E. KEARNS is with the General Electric Company, Erie, Pa.

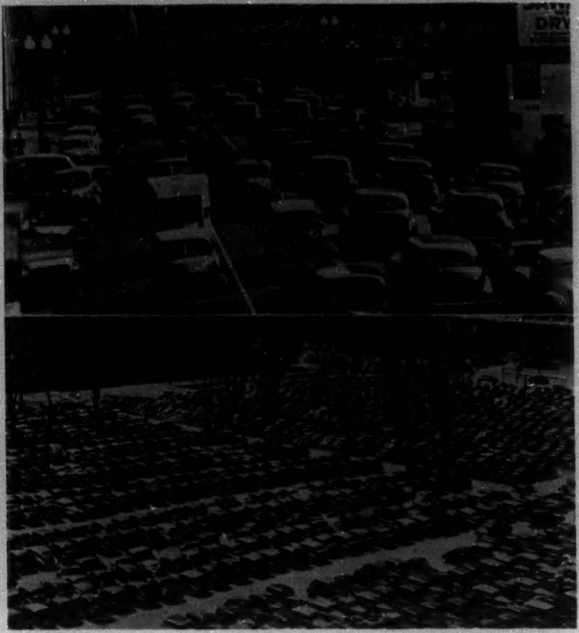


Figure 1. Traffic congestion and parking problems grow steadily worse in large cities

speeding up the movement of people and goods in large urban areas.

Transportation Costs

Viewed from the standpoint of the individual, the whole problem is one of time, convenience, and cost, for all his travels to and from work, recreation, school, and church. Few people stop to think of the cost of traffic congestion in terms of the great delays in the movement of goods. All this lost time and wear and tear on trucks and delivery cars must be paid for by the consumer. Traffic congestion today is costing our country billions of dollars annually. It is estimated that street congestion in New York robs the merchants of over $1,000,000 a day. Boston estimates the cost of trucking delays at over $40,000,000 annually. One authority states that downtown Detroit real estate values have dropped a billion dollars in 15 years, primarily because of the delays and inconvenience of moving in and out of the city.

Moving People

Let us consider a few relationships in the movement of people. One lane on a city street, with cross traffic, can move a maximum of about 1,500 passengers per hour in automobiles assuming the usual average of 1.7 persons per car. On grade-separated freeways the figure is about 2,500 passengers per hour. In comparison where lines converge, a public-transit lane on a city street with buses, trolley coaches, and streetcars can move 9,000 to 13,500 passengers an hour. Asingle-track railway line used for express rapid transit can move as high as 60,000 passengers per hour.

Here is still another way of viewing the problem. Passenger car registrations have increased more than 60 per cent as compared to the period just prior to World War II. About half of the automobiles, approximately 20 million, are owned or operated on city streets. There they travel some 100 billion vehicle miles annually. This horde of vehicles produces about twice as many passenger rides as do the public transit vehicles, yet the latter number only some 87,000 or about 0.5 per cent of the number of private autos. They travel about 3 billion miles or roughly 3 per cent of the mileage traveled by the autos. Based on comparable carrying capacity, the vehicle

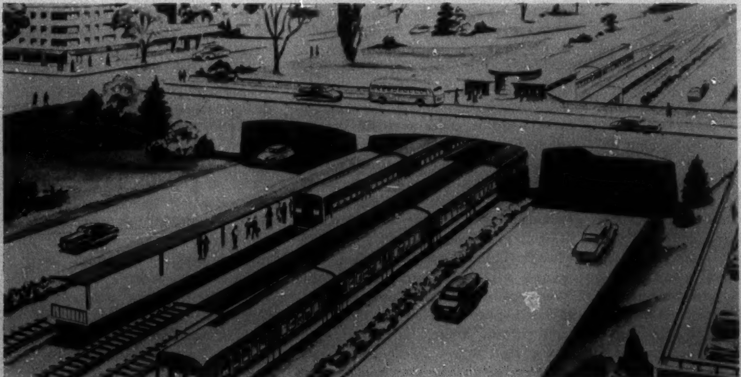


Figure 2. Multiple-unit electric trains multiply the passenger-carrying capacity of a freeway many times

ratio is about 115 to 1 and the mileage ratio about 16 to 1 in favor of public transit. Of equal importance is the fact that public transit vehicles require relatively little parking space for, outside of stopping to load and unload passengers, they keep moving all day long, and when not in use are stored off the streets.

How the Problem May Be Solved

These comparisons obviously lead to the conclusion that a major step in the solution of the traffic congestion problem is to promote greater use of public transit.

This is a simple observation to make, but much careful planning and hard work will be necessary to bring it about. The healthy way to accomplish this objective is to build up public transit systems in our metropolitan areas to a point where they will attract a substantially greater number of riders. In the larger population areas this involves not only public transit but also railroad commuter services operating over a radius of 25 and even 50 miles from the city centers. These services must be modern, clean, attractive, and speedy as judged by today's standards. It is important that the trunk lines for long distance commuting be as free as possible from cross-traffic interference and also from the congestion of vehicle concentration experienced on our most modern freeways and parkways. In large metropolitan areas this can be best accomplished through the use of high-speed rail services completely separated from automotive traffic of all kinds. This network of high-speed trunk lines should be supplemented by a short travel distribution system employing both the private automobile and the modern public transit vehicle in the form of busses, trolley coaches, and streetcars.

Examples of Large Cities

It is most interesting to note that these conclusions are being reached in one large community after another. In July of 1951, the New York Regional Plan Association made public its interesting and thorough studies of the New York commuter problem. The conclusion reached was that development of rail transportation facilities had lagged far behind that for automotive vehicles. Recommendations were made that ways be found to bring the high-speed, rapid transit and commuter developments abreast of those already made for handling rubber-tired vehicles. Further it was recommended that the problem be considered as area-wide and that there be an adequate cooperative effort between the various States, public transit systems, and railroads to bring this greatly needed development into being. Some most excellent considerations and suggestions were voiced in Mr. William Reid's luncheon address at the Sixth Regional Plan Conference in New York City on October 9, 1951. These have been followed by further pertinent addresses and papers by Mr. Reid, Mr. Shoemaker of the Lackawanna Railroad, Mr. Draper of the Long Island Railroad and others, setting forth various aspects of this problem as it pertains to the New York area.

Figure 3. Rush hour traffic floods freeways

Other cities have developed, or are considering, plans for the installation and major expansion of high-speed rail rapid transit. The Boston metropolitan area, through its Metropolitan Transit Authority, has outlined a substantial increase in rapid transit services. It is working toward this end as circumstances and financing permit. Primarily the plan involves using branch line railroad rights of way for its extensions, and envisions financing backed up by general lien credit, that is, financing guaranteed by authorized taxation of real estate in the communities served.

Chicago is expanding its rapid transit system and is endeavoring to demonstrate that this can be accomplished through revenue bond financing; that is, financing backed solely by revenues of the transit operation. Cleveland is starting its rapid transit, utilizing right of way established for that purpose when the Union Terminal project was developed years ago. Financing in this case is also on a revenue bond basis. The city of Toronto is well along with its first rapid transit route, the North-South Yonge Street project. The financing is on a revenue basis but the system, in the final analysis, does have the backing of Toronto general lien security. Active studies of rapid transit are under way at Los Angeles, San Francisco, St. Louis, Detroit, and other large metropolitan areas.

Financing

It is becoming evident that most serious consideration will have to be given to future methods of financing which will have to be employed to raise the large amounts of capital necessary to accomplish the results desired. A considerable school of thought holds the well-justified opinion that these projects, particularly the new rights of way, will have to be financed through general lien credit. This thinking is based on the theory that these rights of way cannot be fully supported out of passenger revenue and, furthermore, that they should not be fully supported out of such revenues because all people in a large metropolitan area benefit greatly by a modern rapid transit system, whether they use the services or not. It makes the movement of people to all points of the area much faster and more convenient. It also contributes in major fashion to traffic congestion relief with consequent reduction in expenses resulting from less traffic delays. In other words, this school takes the position that rights of way for rapid transit should be considered in the same class as other community services, such as streets, water, sewers, and lighting and should be the responsibility of the community at large. Since communities now provide a very considerable share of the streets, parkways, bridges, and tunnels built to handle automotive traffic, it seems logical that, under the same considerations of responsibility, they should provide the rights of way for high-speed rail rapid transit. They contend that the taxpayer's investment per peak passenger handled for the long trip rider is tremendously less for high-speed rail facilities than for paved freeway lanes, however important both may be. This appears to be well founded when one considers their relative capacity. For instance, a double-track rail rapid transit line is equivalent to some 12 to 15 6-lane, grade-separated highways. Moreover, public transit riders do not require major parking areas for their automobiles.

Rapid Transit Rights of Way

There are numerous means of providing rights of way for high-speed, rail rapid transit. No doubt all systems in our large communities will come to use some or all of these. Probably the most expensive is the subway with excavations made specifically for rapid transit. This construction will be required in downtown areas, particularly where no other possible right of way exists. In certain locations the elevated structures will still be desirable. However, the consensus is that this structure reduces the value of business streets. In general, it will be gradually eliminated in order to improve streets by the reduction of noise and the increase of light and fresh air. One of the less expensive methods of obtaining rights of way is to make full use of already existing railroad tracks. This applies both to those now devoted to commuter services and to branch lines which do not now handle capacity traffic. In some instances there are railroad rights of way available which are at present used very little or not at all.

Another much less expensive source of right of way for rail rapid transit is the center mall of a freeway, included in the original design. It is estimated that high-speed rail rapid transit rights of way so provided can be obtained at a cost on the order of one-tenth the cost of subway construction. It is tremendously important that any freeway being installed in our large metropolitan areas be thoroughly studied from the viewpoint of future need of rail rapid transit, and provisions made for it if the studies so indicate. The common experience is that freeways built are never wide enough, so, in the ultimate, if the center mall lanes are not devoted to rail rapid transit they can be devoted to automotive use and their capacities will never be wasted.

A very practical problem immediately poses itself; how to provide the financing for the two or four center mall lanes desired for rapid transit. This must be solved separately from the financing of the automotive freeways which, in most instances, involves not only local but also State and Federal highway funds. Whatever method is used, the financing should be set up to embrace the complete metropolitan area including the commuter's suburban home communities, even though these arrangements may involve several States. The Authority type of organization seems best suited for such a large and diverse undertaking.

Operation

The operation of such lines can be undertaken either by the Authority, or by private enterprise under contract. The city of Philadelphia has an arrangement whereby it provides a certain portion of the rapid transit financing and contracts with the Philadelphia Transportation Company for operation. It has been suggested that private enterprise contracts could carry major incentive provisions to enhance both maximum patronage and greatest efficiency of operation. For example, a certain amount per passenger handled could be given to the private carrier contractor. This would result in the contractor doing all within his power to increase the number of patrons. Further, an additional amount could be given based on the percentage of fixed charges earned out of fares over and above operating costs with specified standards of operation and maintenance, and agreed rates of interest and depreciation.

Excellent studies have been made blocking out desirable rapid transit systems for many of our large communities. Some of these fine engineering recommendations have withered on the vine after their completion, because their importance was not thoroughly heralded to the general public and because the task of forming the necessary organization to design and finance the project looked much too tremendous for anyone to undertake.

Summary

Probably no other community problem is of more importance today than that of public transit. Here is an opportunity for the engineering fraternity to contribute to one of our neediest civic undertakings. As they have so wonderfully contributed to the general advance of our American way of life, they have here another great call for the exercise of their talents; the alleviation of our great traffic congestion dilemma.

Discussion

G. M. Woods (Westinghouse Electric Corporation, East Pittsburgh, Pa.): Mr. Kearns' paper calls attention to a subject which is of vital importance to the continued existence of cities as they are constituted today. As the paper shows, traffic now moves at lower speed in many congested areas than it did in the much maligned horse and buggy days.

Pittsburgh in recent years has benefited by a broad city improvement program. Now transit and traffic are receiving intensive study. The city's deep valleys and steep hills allow relatively few traffic throats from the congested Golden Triangle. The metropolitan area centering there has a population of 2,000,000. Real estate valuations in the central business area have decreased from $325,000,000 in 1930 to $240,000,000 in 1948, or 26 per cent. In the same period, valuations in the county area outside Pittsburgh have increased from $700,000,000 to $970,000,000, or almost 40 per cent.

One leading department store in Pittsburgh has established two branch stores in the suburban area and expects to continue this plan of decentralization. Another leading department store is expanding its facilities in the central area. It is significant that the president of the latter is chairman of the Allegheny County Transit and Traffic Commission whose duty it is to study transit, traffic, and parking conditions and recommend solutions to the problems. Evidentally, he believes that solutions can be found.

In spite of the difficulties generally met in overcoming present intolerable conditions, there is hope in the growing realization on the part of businessmen and city planners that an adequate and efficient public transportation system is essential. The change in the attitude of department store officials is especially noteworthy. Some years ago any restrictions of curb parking received violent opposition. Today the officials of these same stores strongly favor such restrictions because they realize that curb parking space is rarely available when their customers arrive.

It is to be expected that this more enlightened attitude will result in major improvements in public transportation. Continued effort will be required to assure that electrically propelled cars and trolley coaches will be used where they are economically correct.

E. E. Kearns: Every new urban freeway destined to handle substantial intra-city traffic should be designed with at least two lanes of center mall to be devoted exclusively to public transit. Buses, trolley coaches, electric cars, or modern, high-speed, multiple-unit trains, depending upon the number of people to be carried, can be operated over these exclusive rights of way.

A major percentage of people in our larger cities travel by public transit. Therefore, it is only fair that at least two lanes of a modern multilane freeway be devoted to their transportation. Public transit on center-mall lanes multiplies the carrying capacity of a freeway many times. This gives the people of a community maximum transportation service for their tax dollar.

Public transit on such lanes, free from rush hour traffic congestion and cross traffic, can render an attractive, high-speed service. Many people, comparing its advantages with the driving ordeal on congested freeways, and the exasperating search and high cost for auto parking, will prefer to leave their autos outside the highly congested areas. Thus a major contribution will be made toward a solution of the very serious traffic congestion and parking problem.

Precedent is already being set for the allocation of a portion of the local, state, and federal funds appropriated for urban freeway construction to provide center-mall rights of way and other facilities for public transit. Such sound and sensible use of public funds can contribute materially to the very important job of moving people most conveniently and efficiently in our urban areas. Here is an opportunity of the greatest magnitude with far-reaching consequences. The future well-being and convenience of our urban population will be profoundly affected by our embracing or missing this opportunity.

New Series Type Motor for Trolley Coaches

G. M. WOODS
MEMBER AIEE

C. R. STEEN
ASSOCIATE MEMBER AIEE

A NEW trolley coach motor which combines the advantages of the series motor in acceleration and the shunt generator in dynamic braking has been developed. A number of improved features in mechanical and electrical design are included. Particular attention has been paid to obtaining superior commutation both in acceleration and braking.

General

The trolley coach has become established as one of the most important vehicles in city transit. On December 31, 1940, there were only 2,802 trolley coaches in the United States. At the end of 1950, the latest date for which accurate data are available, there were 6,504, an increase of 132 per cent. The position of electrically propelled vehicles in the city transit field is largely dependent on the trolley coach because in the 10-year period the number of surface cars has decreased from 26,630 to 13,800, while the number of subway and elevated cars has decreased slightly. This latter decrease is caused by the scrapping of a number of old elevated cars, the total rapid transit passengers and car miles operated remaining essentially the same. In the 10-year period the number of motor busses increased from 35,000 to 56,820 or 62 per cent.

The rate of trolley coach growth was retarded during the war by the shortage of copper. From the standpoint of financial results, the trolley coach is unrivaled in the transit field. The net income from trolley coach operations in the United States, after all taxes, has been more than $10,000,000 in each of the last three years. City bus operations have had losses of $3,000,000 to $13,000,000 per year in the same period. In many cities having street cars, trolley coaches, and busses only the trolley coaches are operating at a profit. Aside from the more economical operation of trolley coaches, they have marked advantages in comfort, quietness, high schedule speeds, and freedom from odors. The principal deterrent to their use is the necessary investment for overhead lines. The financial status of many transit companies is such that the initial investment required, and not long range economy, is a determining factor in purchases.

Under present conditions, the trolley coach has proved to be the most economical vehicle for the heavier lines in small and medium size cities, as well as for a high percentage of the surface routes in the larger cities including rapid transit feeder lines. In view of the trolley coach's position as the potential successor to the street car as the principal electrically propelled vehicle in city transit, it is imperative that its electrical equipment be outstanding in reliability and low maintenance. The new motor described in this paper and the new control described in a companion paper[1] have been designed with a view to meeting these objectives.

Service Requirements

The modern trolley coach has a seating capacity of 44 to 50 passengers. Its weight varies from 17,000 to 21,000 pounds. The dual rear wheels are driven through double reduction hypoid and spur gears generally having a total reduction of 11.58 to 1. The rear tires have a rolling radius under load of 21.5 inches. The motor must be able to accelerate the coach with average load at a maximum rate of 4 miles per hour per second. The normal rate of dynamic braking is 2 miles per hour per second. The motor armature and commutator must not be damaged by maximum coach speeds of 50 miles per hour. Balancing or free running speed on level roadway with full line voltage applied should be close to 40 miles per hour. The motor design must be such that commutation is successful both when propelling the coach and during dynamic braking, especially from high coach speeds.

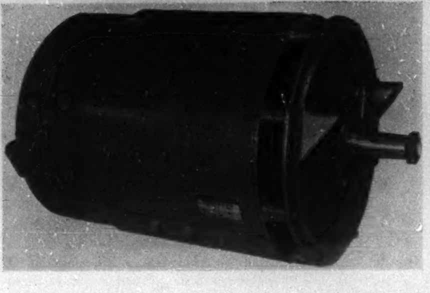
The motor ratings and characteristics have to be suitable for the frequent stop, highly congested service met in cities like Boston and Providence, the high schedule speeds required in Cleveland and Chicago, and the severe grades approaching 20 per cent in San Francisco and Seattle. Only by standardization on one motor for the varied conditions of operation can the cost be held to a reasonable value, although the motor speed and rating, or both, may be higher than required for some applications.

Design Objectives

The series type traction motor, without question, provides the most desirable speed and torque performance for transportation vehicles. However, the series motor is difficult to control for dynamic braking and leaves something to be desired as far as smoothness of braking is concerned. Paralleling the development of trolley coaches, the desire has increased for smooth dynamic braking performance as well as for a reduction in the complexity of the control and in the duty imposed upon its working parts. These requirements can best be met by co-ordinating the designs of the control and the traction motor. Thus, the type *1442-N1* motor has been designed to complement the new greatly simplified trolley coach control.

Paper 52-37, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted October 23, 1951; made available for printing December 5, 1951.

G. M. WOODS and C. R. STEEN are with the Westinghouse Electric Corporation, East Pittsburgh, Pa.



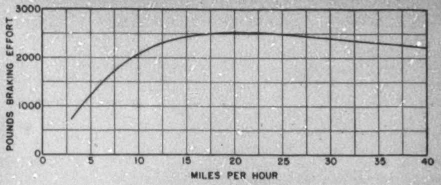


Figure 1 (left). Type 1442-N1 140-horsepower 600-volt 1,700-rpm Super-Series trolley coach motor

Figure 3 (above). Dynamic braking curve of type 1442-N1 motor, 11.5:8 to 1 gear ratio, 43-inch wheels

The design of the new trolley coach motor has been based on two principal objectives:

1. A motor having the same good accelerating performance as the previous type *1442-A* series motor including comparable flashing resistance.
2. A motor providing a smooth and essentially constant rate of dynamic braking without requiring control notching as is the case with the series type of machine.

These aims have been met with the type *1442-N1* traction motor, shown in Figure 1, which has the same mounting dimensions as the earlier series motor. This motor has a 1-hour rating of 140 horsepower, 194 amperes, 1,700 rpm at 600 volts self-ventilated.

The desirable motoring performance shown on Figure 2 is obtained by use of the series field only, thus retaining the inherent stability against flashing that is characteristic of the series type motor.

Smooth, notchless, and almost constant dynamic braking effort is provided over a wide speed range, as shown in Figure 3, by adding a shunt field winding which is used with a differential control circuit. The extra space required by the shunt winding has been obtained by increasing the frame diameter by $1^1/_2$ inches. Figure 4 shows the schematic diagram of the circuit employed in order to accomplish this automatic braking performance. It will be noted that both the shunt field current which is supplied from the line and the current generated by the armature must pass through the braking resistor, R_b. At high vehicle speeds, the armature current is large and produces a large voltage across the braking resistor, thus reducing the voltage available across the shunt fields which decreases the shunt field current and limits the generator torque. Conversely, at low speeds the armature current and braking resistor voltage are small thereby increasing the shunt field excitation and essentially maintaining constant torque.

Mechanical Design Features

The long section of the new Super-Series motor, Figure 5, shows many of the important mechanical features that have been incorporated in its design.

Mechanically, the design centers around the integral frame construction with its welded-on commutator-end housing. This type of construction assures correct bearing and brush-holder alignments since the commutator-end housing is machined as part of the frame, and being permanently attached, this alignment will not be altered.

The brushholders are the so-called end-mounted type and are clamped in blocks that are welded to the machined inside face of the frame. The integral type frame construction makes possible the precise location of these brushholder blocks by means of a fixture that locates directly from the commutating pole bolt holes.

The brushholders are precision cast by the permanent mould process which results in accuracy and uniformity. The spring saddle and tension key construction provides for easy brush pressure adjustment. A factor of value to transit maintenance departments is that all four brushholders are the same, thereby requiring only one item for stock.

An all-out effort has been made to provide an exactness of brush position that will be permanent. In addition to the aforementioned steps that have been taken to assure accurate components, there is an inherent advantage to the end-mounted brushholder. It will be noted that the clamping action on the brushholder pins is only slightly farther from the center of the armature shaft than is the contact of the brushes with the commutator. As a result, a slight tilting of the brushholder pins in the mounting block grooves will produce negligible movement of the brush position, thereby maintaining good brush spacing. This is a distinct advantage over the conventional side-mounted holder where the clamping action takes place a great distance from the brush contact with the commutator.

Another feature that is of considerable importance from the maintenance standpoint is found in the drive-end housing and bearing assembly construction. The

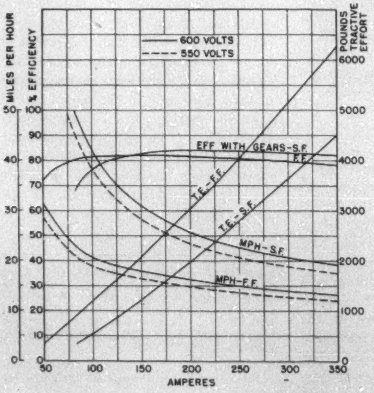


Figure 2. Characteristic curve of type 1442-N1 motor, 600 volts, 11.58 to 1 gear ratio, 43-inch wheels

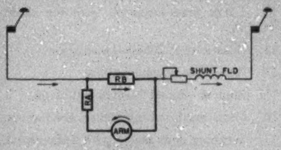


Figure 4. Schematic diagram of dynamic braking circuit

drive-end bearing assembly has been redesigned to permit the removal of the housing and even the armature fan without requiring that the bearing inner race and rotating grease throwers be pulled from the armature shaft. This has been accomplished without sacrificing the effectiveness of the labyrinth seals which have been proved to be so effective.

Years of experience gained from operation of earlier trolley coach equipments has pointed out several minor mechanical changes that were desirable. These modifications, as well as applicable developments in materials and processes that have occurred in the meantime, have been incorporated in the design of the Super-Series motor. Some of these items are worthy of mention in that they are related to improved performance, accessibility or convenience, or directly to reduced maintenance labor.

1. Machined commutator cover seat for tight fit of covers.
2. Reversible commutator covers permitting latches to be located on either side of the motor.
3. End-mounted brushholders providing accessibility in the commutator compartment that is impossible with the side-mounted type of holders.
4. Full reaction type brushes for best commutator riding qualities.
5. Synthetic bonded bar and vee-ring mica adopted for the commutator, increasing its strength and stability at higher temperatures.
6. High melting point solder adopted for soldering armature conductors in the commutator risers and for the armature bands.
7. Flash ring added to the commutator-end bearing cartridge to protect the bearings against damage from flashing current.
8. Brushholder cross-connections relocated in the wiring-around-frame on the field coil side to provide full access to the commutator vee-ring string band for cleaning and painting.
9. Through bolts used to fasten the drive-end housing to the frame, making possible the use of a socket or spinner type wrench.

Electrical Design Features

The principal electrical change was made necessary by the addition of the shunt field winding for dynamic braking purposes. The main field coil which contains both the series and the shunt windings is of unitized construction. The series winding is wound from large rugged copper strap which is then insulated overall as a unit. The shunt section is wound with a sufficiently large round wire to permit winding in turns and layers so that each wire is in its proper position, thus avoiding laps and cross-overs and the possibility of short-circuits developing between turns. The shunt section is also completely insulated as a unit and is then assembled with the series coil and the entire coil is taped over-all resulting in a single mummified main field coil unit.

Both the main and commutating field coils have been insulated with Silastic tape which is a Silicone-rubber-impregnated glass tape. Silastic tape is a fully class *H* insulating material, but due to its rubberlike characteristics, it has other desirable properties. It provides an excellent seal against creepage due to the presence of oil or water since it has considerable flexibility in taping and the Silastic has a tendency to vulcanize between layers during baking in a manner similar to rubber. Silastic insulation will also have exceptionally long life when used within normal class *B* temperature rises as have been maintained in this design.

Two improvements have been made in the armature winding, which, incidentally will not affect its interchangeability with the earlier series type motors.

1. The armature conductors have been changed to have double-glass insulation which represents a considerable improvement over the previously used conductor insulations consisting of either asbestos or enamel-asbestos. The glass insulation is nonhydroscopic in nature and great strides have been made in its application to the wire to the extent that the wire can be bent around a normal forming pin without cracking or peeling as is common with asbestos insulated wire.
2. The wrapper ends have been sealed well onto the diamond portion of the coil with Teflon tape. Teflon is a valuable new class *H* insulating material derived from the fluoro-carbon family that has dielectric strength even higher than that of mica, but, in addition, has the advantage of having a considerable amount of flexibility. It is this latter quality that makes Teflon so valuable as a wrapper sealer since it tapes smoothly around the bend of the coil without exhibiting the voids that make mica tape so susceptible to creepage breakdown.

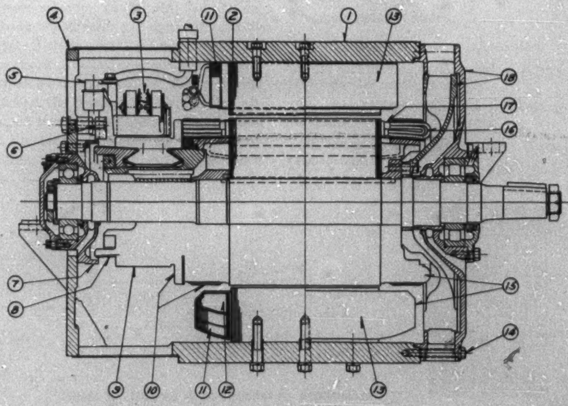
Both armature and field windings, in addition to the normal varnish treatments, receive a spray coat of red insulating enamel. The enamel produces a hard slick surface that is effective in shedding moisture and also resists the tendency for dirt to accumulate.

Resistance to Flashing

Since the trolley coach receives its power from an overhead line, the traction motor is subjected to voltage interruptions and surges that are conducive to flashing at the commutator. The straight series motor was unexcelled with regards to its resistance to flashover, and in the design of the Super-Series motor, the utmost has been done to maintain this good flashing characteristic.

However, the addition of the shunt field winding has a deleterious effect upon the resistance of the motor to flashing. If

Figure 5. Type 1442-N1 motor—longitudinal section



the shunt winding could be left open during motoring, it would have no effect upon the series motor performance. However, since the shunt winding encompasses the main field pole, it is coupled by mutual inductance to the series field winding. Thus, whenever there is an interruption or surge of current in the motor circuit that causes a change in the main pole flux, a voltage is induced in the shunt coils. Since there are a large number of turns in the shunt winding, the induced voltage would be of a magnitude that would exceed the safe dielectric limits of the coil insulation. The remedy for this excessive induced voltage is simple, but it is not accomplished without a price. A field discharge resistor is connected across the shunt field leads so that, during the surge, it permits current to flow in the shunt coil. This current produces ampere-turns in the shunt winding that oppose the build-up or decay of flux in the main pole, and since the induced voltage is proportional to the rate of change of the flux, it is thereby reduced to a safe value as predetermined by the resistance of the discharge resistor. However, this action in reducing the rate of change of the motor flux constitutes additional damping action and reduces the ability of the motor to withstand power interruptions.

In the design of the Super-Series motor, the loss in flashing resistance resulting from shunt field damping action, has been offset by the development of laminated commutating poles. Interruption tests taken with the Super-Series motor confirm that its ability to withstand power interruptions or surges is essentially the same as for the older series motor.

Design Features

1. Rigid one-piece frame and commutator housing.
2. Laminated commutating poles.
3. Easy brush tension adjustment.
4. Machined commutator cover fit.
5. Precision permanent mould brushholders.
6. Accurate brushholder mounting.
7. Flash ring protection.
8. Full access to string band.
9. Synthetic bonded mica for stable commutator.
10. High melting point solder.
11. Heavy strap for series and commutating pole windings.
12. Large glass-insulated wire for shunt winding.
13. Silastic-tape field coil insulation.
14. Accessible housing bolts.
15. Smooth enamel finish on windings.
16. Glass-insulated armature conductors.
17. Wrapper ends sealed with Teflon tape.
18. Housing and armature fan removable without pulling bearing assembly from shaft.

Conclusion

Five sample trolley coach equipments of this type have been in service in Boston and Chicago since the autumn of 1950. Their performance from the standpoint of the coach operators has been most successful. Experience shows that the objectives of the design have been met. The motor commutator surfaces are clean and have an excellent polish, confirming that the good commutating characteristics of the series motor have been retained.

Reference

1. A NEW CONTROL FOR TROLLEY COACHES, N. H. Willby. *AIEE Transactions*, volume 71, part II, 1952, pages 63–65.

No Discussion

A New Control for Trolley Coaches

N. H. WILLBY
MEMBER AIEE

Synopsis: A new control system has been developed to operate with the new trolley coach motor described in a paper by Woods and Steen.[1] In keeping with the trolley coach reputation as the lowest cost vehicle, this control has been designed for maximum reliability and minimum maintenance.

THE NEW trolley coach motor with its series field for acceleration and its shunt field for braking is admirably suited to help the control designer obtain the objective of reliability and low maintenance. In acceleration, it operates as a simple series motor. To control such a motor, we need only to commutate a resistor automatically in the well-known manner. In braking, the motor operates as a line excited shunt generator whose excitation is modified by the armature current. This produces a 1-step dynamic brake, giving a practically constant dynamic braking rate from maximum speed down to about 10 miles per hour with a declining rate thereafter. This 1-step system means that only a dynamic brake setup circuit is required, and no control functioning is necessary during the dynamic brake cycle.

Control Scheme—Acceleration

The control scheme for acceleration is shown in Figure 1. It shows the series motor with line switch, seven resistor switches, and two field shunting switches. A current limit relay measures the current for automatic acceleration and an overload relay protects against excessive current. After all resistance is cut out, speed is further increased by field shunting.

Control Scheme—Braking

Figure 2 shows the circuits during dynamic braking. The armature is closed on a resistor which dissipates the braking energy. The shunt field is energized from the line, and its circuit passes through a portion of the dynamic brake resistor. The relative polarity is selected so that increase in armature current results in a drop in shunt field current. By this means, a self-regulatory effect is obtained which avoids the necessity of any control functioning during dynamic braking.

Different rates of dynamic brake are obtained by varying the amount of resistance in series with the shunt field circuit.

One of the elements of the design is the proper selection of the *RA* and *RB* resistors which form the dynamic braking loop. A change in the *RA* resistor changes the speed at which the maximum retarding effort is obtained. A change in *RB* changes the amount of braking effort obtained at all normal speeds. By selection of the proper combination of these resistors, a braking effort curve is obtained that is essentially constant from maximum down to ten mph. At still lower speed, the braking effort fades out as standstill is approached. Figure 3 shows shunt field current, armature current, and retarding effort during the braking cycle.

This scheme of braking is well suited to trolley coach requirements. The absence of any control sequence during braking keeps the functioning of control parts to a minimum and reduces maintenance. The simplicity of the scheme insures a reliability that cannot be equalled by schemes requiring more complicated control functioning. Braking response is immediate and uniform, and this feature is always appreciated by operators.

Selection of Apparatus

After a given scheme of control has been chosen, the next step is to select from existing designs or make new designs of components which will best perform these functions. It also is necessary to select the form of energy used for control purposes and finally assemble these components into suitable panels. This new trolley coach control has been assembled into:

Main control panel, Figure 4.
One pair of resistor panels, Figure 5.

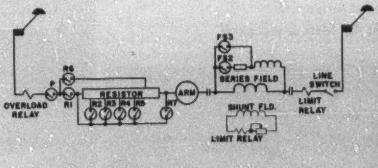
The main panel contains:

One large contactor to act as a line switch.

Master and brake controller with a current limit relay for acceleration only.

Cam controller to commutate accelerating resistance.

Four small contactors to regulate shunt field current.



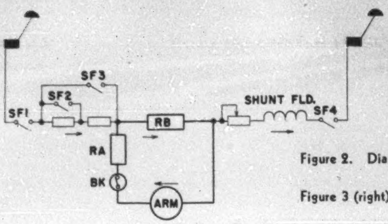
ACCELERATING SEQUENCE CHART

POS.	LS	P	R1	R2	R3	R4	R5	R6	R7	FS2	FS3
OFF			●								
1	●	●	●								
2	●	●	●	●							
3	●	●	●	●	●						
4	●	●	●		●	●					
5	●	●	●			●	●				
6	●	●	●				●	●			
7	●	●						●	●		
8	●	●		●				●	●		
9	●	●		●	●			●	●		
10	●	●		●	●	●		●	●		
11	●	●		●	●	●	●	●	●		
12	●	●	●	●	●	●	●	●	●		
13	●	●	●	●	●	●	●	●	●	●	
14	●	●	●		●	●	●	●	●	●	●

Figure 1. Diagram of connections for acceleration

Paper 52-38, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted October 24, 1951; made available for printing December 13, 1951.

N. H. Willby is with Westinghouse Electric Corporation, East Pittsburgh, Pa.

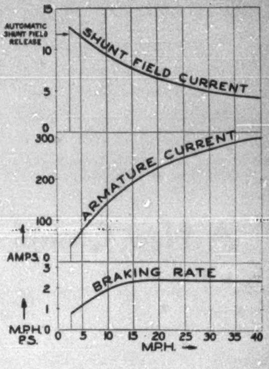


BRAKING SEQUENCE CHART

POS.	BK	SF1	SF2	SF3	SF4
OFF	●				
1	●	●			●
2	●	●	●		●
3	●	●	●	●	●

Figure 2. Diagram of connections for braking

Figure 3 (right). Characteristic braking curves



Overload relay.

Relay to open the shunt field at the end of the braking cycle.

Silicon carbide disk connected across the shunt field to reduce peak voltages.

Master and Brake Controller

The master and brake controller contains three shafts which are connected by a linkage to the operating pedals. One shaft controls acceleration, another braking, and the third reverses the main motor field to reverse direction of motion. The acceleration shaft also is tied through a cam to the spring which calibrates the current limit relay. By this means, a wide range of accelerating rates is available to the operator, giving the flexibility so desirable in city transit service.

Power Brake Changeover

The cam controller, whose major function is to commutate the accelerating resistance, has built into it a power brake changeover switch. This device includes two cam switches actuated by a separate air cylinder and magnet valve. One cam switch closes the power circuit and the other closes the dynamic braking loop. Both switches are operated by a single cam. This device eliminates all electrical interlocking between power and brake circuits. It is spring-returned to the brake position.

Control of Notching

A current limit relay is the customary method of obtaining automatic acceleration on transit vehicles. After current has built up to the indicated value, this device will allow each successive notch to be taken at the same current value. Another factor to be considered is the natural time of the control sequence. This governs the initial build-up of tractive effort and also affects the time of power response when power is re-applied. Normally, the initial build-up of tractive effort on transit vehicles reaches 4 miles per hour per second at the end of the first second. This rate of build-up should be as high as practical for maximum schedule speed but is limited by the tolerance of passenger comfort.

Tests have shown that passenger comfort is a function not only of the rate of build-up of acceleration but also of the actual rate of acceleration itself. As the acceleration approaches a high value, passenger comfort is improved if further increases of acceleration are taken more slowly. It is particularly objectionable if one more notch is taken just as the maximum current relay setting is reached.

On this equipment, a feature has been added to modify the customary build-up of accelerating tractive effort. A small coil has been added to the current limit relay, and this coil has been connected across the shunt field of the main traction motor. This field is not used during the accelerating period but, of course, reflects by transformer action any current increase in the series field. Thus when each notch is taken, the limit relay receives an impulse tending to open it. When main current approaches the current limit relay setting, this added impulse opens the relay, slows down the rate of progression and rounds off the tractive effort curve as shown in Figure 6. The amount of this rounding effect can be adjusted by a resistor in series with the coil. Figure 7 shows a similar notching curve without this feature.

This is a simple device with no added moving parts and no added electrical contacts. It also serves to positively open the current limit relay on each succeeding notch, thus preventing any chance of skipping notches. Various positive notching schemes are common in the art but involve moving parts and electrical contacts. The method of accomplishing this function described here is in keeping with the purpose of the over-all design of maintaining a high standard of reliability in the trolley coach.

Conclusion

At the time this paper was written, a small number of these new trolley coach equipments have been in service about a

Figure 4. Main control panel

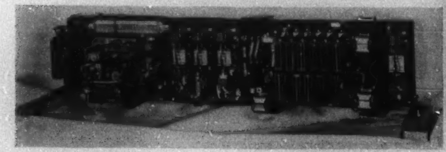
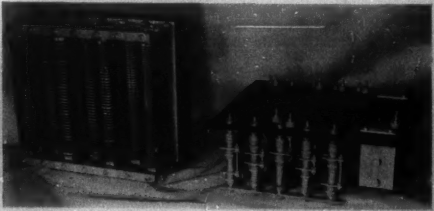


Figure 5. Resistor panel



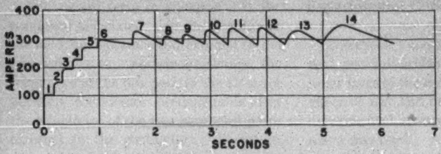


Figure 6. Notching curve with limit relay notching coil

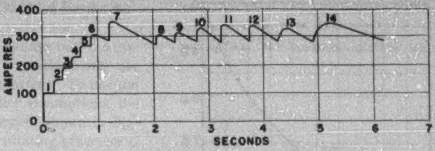


Figure 7. Notching curve without limit relay notching coil

year in two different cities. They have fully lived up to expectations for reliability, low maintenance, and operator approval. Two larger groups of these equipments are just going into service in Cleveland and Boston. Of these, about 30 coaches now are running in revenue service in Cleveland. Their record has been fully equal to the small number of equipments put in service a year ago.

Reference

1. New Series Type Motor for Trolley Coaches, G. M. Woods, C. R. Steen. *AIEE Transactions*, volume 71, part II, 1952, pages 59–62.

Discussion

H. G. Moore (General Electric Company, Erie, Pa.): The authors of this paper and a companion paper[1] are to be commended on their recognition of the advantages of shunt excitation of the motor in providing simple and promptly responsive dynamic braking for trolley coach service. The General Electric Company adopted this principle in 1938 following successful demonstration of its practicability some years earlier in Europe. More than 4,400 equipments having the dynamic braking characteristics described in these papers have been supplied in the ensuing 13 or 14 years.

In the General Electric equipments full advantage is taken of the presence of a shunt field by using it in acceleration as well as braking. This provides several important features as follows:

1. The motor excitation is increased, resulting in lower line current for a given rate of acceleration and reduced losses in accelerating resistors.
2. A means of speed control by regenerative braking is provided at all speeds above approximately 18 miles per hour by the simple use of the power pedal to control shunt field excitation.
3. Shunting of the series field with the attendant use of inductive shunt and additional power switches and resistors is made unnecessary. In either case the motor operates as a simple series-excited motor when accelerating at speeds above approximately 20 miles per hour.

Reference

1. See reference 1 of the paper.

E. E. Kearns (General Electric Company Erie, Pa.): In connection with these new electrical developments for trolley coaches, it is interesting to note that since its inception in the early 1930's, the trolley coach fleet has been developed to something over 9,000 vehicles in the United States and Canada. This vehicle has a number of major advantages which justify its continued expansion in serving the transit industry.

Its operating costs are comparatively low. Maintenance of the electric equipments is substantially less than for a self-contained power plant. Furthermore, electrical braking increases brakeshoe life some three to five times. The large amount of power available from the unlimited trolley supply permits faster schedule speeds. The vehicle, overhead, and substations have exceptionally long life and consequent low fixed charges.

From the revenue attracting viewpoint, the trolley coach also has characteristics which enhance its financial showing. Its freedom from fumes, lack of noise and vibration, and fast smooth performance are greatly appreciated by transit patrons, resulting in maximum patronage and revenues.

Although the economic considerations pertaining to each community and each individual line must be analyzed in detail, in general the trolley coach field of economic superiority can be defined in general terms. It exists where large-capacity, rubber-tired units are needed to supply from 1 to 10 minute headways in rush hours; where passengers-per-hour capacities in rush hours range from 400 to 4,000 at the point and in the direction of heaviest travel; where millions of annual seat-miles per route mile range from 1 to 20; or where vehicles per mile of 2-way route range from one to ten. Wherever the service demands of established local stop trunk lines are in the above categories, serious study should be given to the advantages and benefits of trolley coaches whenever modernization plans involve purchases of new vehicles.

In addition to the 9,000 trolley coaches now in service or on order in the United States and Canada, it is estimated that some 25,000 more such vehicles are economically warranted on the various transit properties of this continent.

E L. Mueller (Genera Electric Company, Erie, Pa.): The authors of this paper and a companion paper[1] are to be congratulated on what appears to be considerable progress.

In the long run, we shall all benefit.

The important thing is that the trolley coach should make the best showing against the motor bus. This is particularly true as regards maintenance expense. For this reason many cities, large and small, continue to build large fleets of trolley coaches as the backbone of their surface systems. Motor buses are used to advantage in the lower density services.

The maintenance requirement of a trolley coach is phenomenally low.

American Transit Association averages last year indicated maintenance and servicing expenses averaged $2,500 per vehicle per year for trolley coaches. This included overhead wire maintenance. The corresponding expense was $3,500 per vehicle per year for smaller motor buses working in lighter service. It has been the difference of $1,000 per vehicle per year that has counted.

We are glad to acknowledge pending progress for trolley coaches.

Reference

1. See reference 1 of the paper.

N. H. Willby: As Mr. Moore states in his discussion, a large number of motors using shunt field excitation in dynamic braking have been in successful operation in this country for many years. However, these motors operate with cumulative compound fields in acceleration and differential compound fields in braking which make them more susceptible to poor commutation and motor flashing. Mr. Moore claims lower accelerating current and lower resistor losses for the compound motor. Actually this is determined by design and not by the field connections. The series connected motor used with the equipment described in this paper produces normal accelerating tractive efforts at a lower current than the compound motor referred to by Mr. Moore. This advantage of the series motor increases with heavier duty.

Speed control with the compound motor above 18 miles per hour has some theoretical value, but it is little used because this speed is too high for congested areas where speed control is most needed. The actual use of regenerative braking has been found to be negligible both in this country and in Europe. Regeneration also introduces the problem of overvoltages and reverse currents at rectifier and synchronous converter substations.

Series field shunting with inductive shunts and necessary switching are required for the acceleration of the series motor. The addition of this small amount of equipment is fully justified, by the superior performance obtained in acceleration by the series motor.

Bus Bar Design for High-Frequency Induction Heating

J. W. WILLIAMSON
MEMBER AIEE

WHERE motor-generator sets are used, induction heating coils are frequently located at points remote from the high-frequency generator in a manufacturing plant. In such cases, one or more transmission lines are required. These may be either of the cable or rectangular bus bar type. A subsequent paper will consider the use of cables. The present paper deals with rectangular bus bars. Primary attention is given to the range of frequencies commonly generated by inductor alternators. Current carrying capacity and resistance and reactance per unit length are tabulated for 10,000, 3,000, and 1,000 cycles. Equations are given which apply to much higher and lower frequencies as well. The information supplied has not previously been available to users of induction heating equipment.

Resistance and Reactance

A typical arrangement of rectangular bus bars is shown in Figure 1, with reference to which the combined resistance of the two bus bars and the reactance, both per unit length of run (dimension L) are:

$$R_1=\frac{2\rho}{Wd}F \tag{1}$$

$$X_1=\frac{2\rho}{Wd}G+2\pi f\mu_0\frac{S}{W}J \tag{2}$$

where

W = width of either bus bar
S = spacing between bus bars
d = reference depth for conductor material at given frequency $=\sqrt{\rho/(\pi f\mu_0)}$
μ_0 = permeability of air or conductor material, expressed in $\dfrac{\text{(appropriate lines)/(unit area)}}{\text{(unit current)/(unit length)}}$

F and G are functions of the ratio of conductor thickness to reference depth d, and are plotted in Figure 2
J depends on the ratio S/W of spacing between conductors to width of either conductor and is plotted in Figure 3
ρ = resistivity of conductor material
f = frequency, cycles per second

Equations 1 and 2 apply in any consistent units. However, for the remainder of this paper it will be assumed that R_1 and X_1 are to be expressed in ohms per inch of run. Then, for copper at 85 degrees centigrade, $\rho=0.85\times10^{-6}$ (ohm inches).

With reference to the definition of μ_0, the meaning of "appropriate" is implied in the statement that a rate of change of one appropriate line per second produces one unit of potential difference in whatever dimension system is used. In the practical system of electrical units potential differences are expressed in volts. Therefore, for the equations here presented to apply the appropriate flux line is 10^8 electromagnetic system lines = 1 weber. In the units to be employed

$$\mu_0=3.20\times10^{8}\ \frac{\text{webers per square inch}}{\text{amperes per inch}}$$

and the reference depth d is numerically given by

$$d=3\,160\sqrt{\rho/f}\ \text{(inches)}$$

Considering copper at a temperature of 85 degrees centigrade

$d=0.029$ inch for a frequency of 10,000 cycles per second
$=0.052$ inch for a frequency of 3,000 cycles per second
$=0.092$ inch for a frequency of 1,000 cycles per second

To understand equations 1 and 2, first suppose that the reference depth d, which is a measure of how far from the inner faces the current penetrates if the conductors are sufficiently thick, is small in relation to the thickness T. (Wd), the product of the conductor width and reference depth, may be thought of as the effective area through which current flows in each conductor. Therefore, $\rho/(Wd)$ is the resistance per unit length for either conductor, and $2\rho/(Wd)$ the resistance unit length of run. It turns out that $2\rho/(Wd)$ is also the reactance per unit length of run, associated with lines of flux within the conductor material. The factors F and G in equations 1 and 2 are practically equal to one providing the ratio of thickness to reference depth is 1.2 or greater. For copper bus bars, this includes all cases where the thickness is at least 1/8 inch and the frequency is at least 1,000 cycles. For smaller values of T/d, F and G can be obtained from Figure 2.

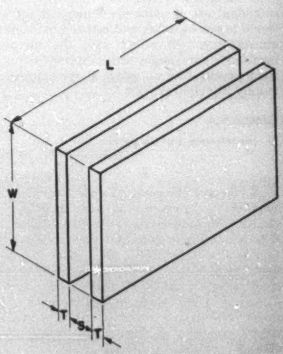


Figure 1. Typical bus bar arrangement

Paper 52-68, recommended by the AIEE Electric Heating Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted October 24, 1951; made available for printing December 14, 1951.

J. W. WILLIAMSON is with the Ohio Crankshaft Company, Cleveland, Ohio.

The second term in equation 2 is the reactance associated with lines of flux in the air space between and external to the two conductors. For close spacings, the reluctance of the flux path in the air space between conductors predominates over the reluctance of the two paralleled paths external to the conductors. Therefore, where I_{inst} is the instantaneous current (amperes) carried by either conductor, the magnetizing force (amperes per inch) between conductors in I_{inst}/W. The magnetic flux density (webers per square inch) produced by this magnetizing force is $\mu_0 I_{\text{inst}}/W$. Multiplying by the spacing S gives $\mu_0 I_{\text{inst}}S/W$, which is the number of lines of flux (webers) per inch of run. The corresponding components of inductance and reactance per inch of run are respectively $\mu_0 S/W$ and $2\pi f\mu_0 S/W$. As the separation is increased, the flux density between conductors becomes less than that which would exist if the reluctance of the air space external to the conductors were negligible in relation to the reluctance of the flux path between conductors, and the correction factor J, see Figure 3, is required in the second term of equation 2. This correction factor was derived from "Formulas and Tables for Calculation of Mutual Self Inducance"[1]

Impedance values calculated from equations 1 and 2 for typical widths and spacings and frequencies of 10,000, 3,000, and 1,000 cycles are given in Table II.

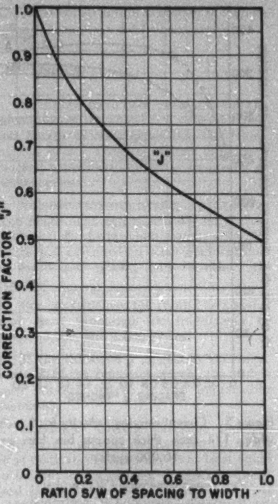


Figure 3. Correction factor J

Experimental Checking of Formulas

Equations 1 and 2 assume that the current per unit width is constant over the inner faces of the conductors. To test the effect of this and other assumptions, the experimental arrangement shown in Figure 4 was used. A rotating generator G supplied 10,000-cycle power. This power was metered through ammeter $A1$, wattmeter $W1$, varmeter $R1$ and voltmeter $V1$. The capacitor C was adjusted for approximately unity power factor at these meters, to obtain accurate power readings. A standard type of induction heating transformer T was used to step down voltage and increase current. The bus bars being tested were connected to the fishtails of this transformer and spread out near these fishtails to permit the insertion of a current transformer, $CT2$. The impedance over the run L was found by dividing the voltage between points $P1$—$P2$ by the current metered through current transformer $CT2$. To determine losses over the length L two readings were taken through wattmeter $W1$, both with the same bus bar current. One of these readings was taken with this current flowing over this entire length. The other reading was taken with a jumper inserted at $P1$—$P2$. Subtracting and correcting for differences in capacitor and transformer losses gave the watts loss over length L. Dividing this loss by the current squared gave the resistance. Having found the impedance, the reactance could then be determined. In most cases the reactance is practically equal to the impedance. The length L was 24 inches in all cases.

Table I gives a comparison between experimental and calculated impedance values for several spacings of 2- and 4-inch wide bus bars. The frequency was 10,000 cycles. For close spacings equations 1 and 2 give accurate results, as might be expected since here the current per-unit width should be constant over the inner faces of the two conductors. If the spacing is increased, the current per unit width along the inner faces becomes less uniform, and there is some current along the side faces or edges. For 1/16-inch thick bus bars the effect of current carried by the edges is minor, and the resistance per unit length increases with increased

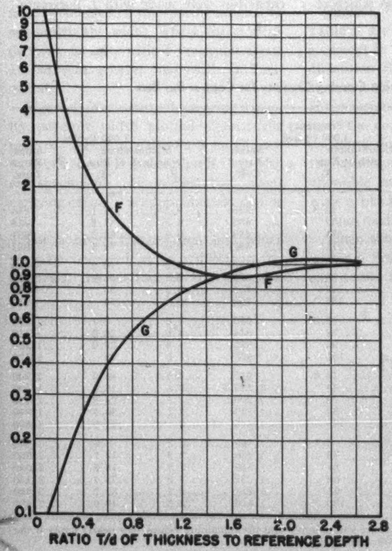


Figure 2 (left). Correction factors F and G

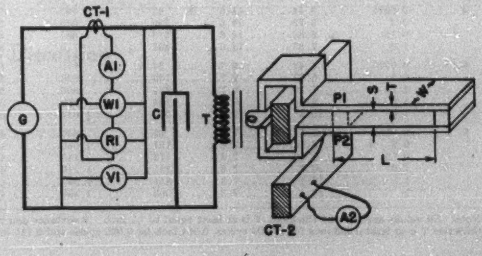


Figure 4 (below). Experimental arrangement used to determine impedance and current carrying capacity

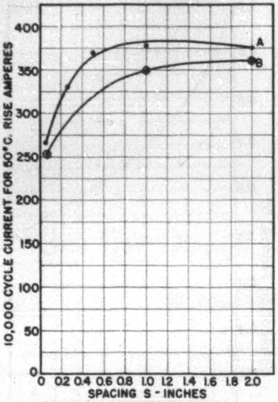


Figure 5. Current carrying capacity for 2 inch wide, 1/4 inch thick copper bus bars on 10,000 cycles

A—2-inch dimension (W) vertical
B—2-inch dimension (W) horizontal

spacing, due to uneven current distribution on the inner faces. For 1/4-inch thick bus bars sufficient additional current is carried by the edges with increased spacings so that the resistance per unit length decreases. High frequency currents tend to distribute themselves along the surfaces of the conductors carrying them in such a way as to minimize reactance. Therefore, where the spacing is appreciable and the current density is not uniform over the inner faces and zero elsewhere, as assumed in calculations, the actual reactances are less than the calculated values. The reduced accuracy of equations 1 and 2 when the spacing is more than about one-fourth the width is not a serious limitation on their usefulness. As will be described, large spacings do not increase current carrying capacity, and in general would be used only to avoid danger of voltage breakdown.

Table I. Comparison Between Calculated and Measured Resistance and Reactance Per Unit Length for Frequency of 10,000 Cycles

Width W, Inches	Thickness T, Inches	Spacing S, Inches	Calculated Resistance	Micro-ohms Per Inch of Run Calculated Reactance	Measured Resistance	Measured Reactance
2	0.0625	0.064	29.2	90.3	28.2	89.4
		0.128	29.2	146	29.2	148
		0.256	29.2	248	29.2	230
		0.510	29.2	421	28.2	386
		1.00	29.2	682	30.2	612
		1.97	29.2	1022	30.5	924
2	0.25	0.064	29.2	90.3	28.8	89.4
		0.510	29.2	421	27.8	371
		1.00	29.2	682	26.0	587
		1.97	29.2	1022	26.0	891
4	0.25	0.255	14.6	131		127
		0.510	14.6	233	15.5	216
		1.00	14.6	401	15.0	366
		2.00	14.6	669		608

Note: Experimental values are corrected to temperature of 85 degrees centigrade so as to be comparable with calculated values.

Current Carrying Capacity

The current which a transmission line can carry is here taken as that which results in a temperature rise of 50 degrees centigrade above a 35-degree centigrade ambient. The amount of power per unit length which a pair of rectangular bus bars can dissipate in still air with a given temperature rise depends on the following:

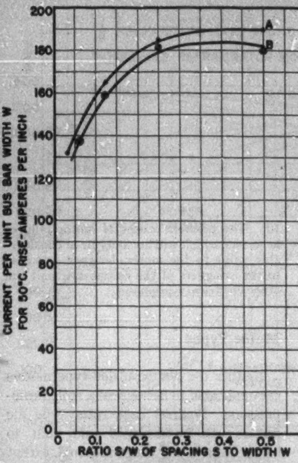
1. The spacing.
2. Whether the dimension *L* (see Figure 1) is horizontal or vertical.
3. If *L* is horizontal, whether *W* is horizontal or vertical.
4. The color of the surface, which determines how much heat can be dissipated per unit time by radiation.

All tests to determine current carrying capacities were made with *L* horizontal,

Table II. Resistance and Reactance Per Unit Length and Current Carrying Capacity for Copper Bus Bars

			10,000 Cycles			Frequency 3,000 Cycles			1,000 Cycles		
Width W, Inches	Spacing S, Inches	Watts Loss per Inch of Run for 50°C Rise	Michro-ohms per Inch of Run R_1	X_1	Amps. Rated Curr.	Michro-ohms per Inch of Run R_1	X_1	Amps. Rated Curr.	Michro-ohms per Inch of Run R_1	X_1	Amps. Rated Curr.
2	0.0625	1.81	29	88	250	16.0	23.8	338	9.2	15.1	444
	0.5	3.45	29	415	345	16.0	132	466	9.2	47.8	613
	0.75	3.67	29	557	355	16.0	174	480	9.2	62.0	631
	1.0	3.67	29	682	355	16.0	211	480	9.2	74.5	631
3	0.0625	2.52	19.5	60	360	10.7	22.7	487	6.2	10.2	639
	0.5	4.88	19.5	294	500	10.7	93	676	6.2	33.6	888
	0.75	5.48	19.5	406	530	10.7	127	716	6.2	45	941
	1.0	5.58	19.5	501	535	10.7	155	723	6.2	54	950
4	0.0625	3.44	14.6	45	485	8.0	17.2	655	4.6	7.7	861
	0.5	5.72	14.6	229	625	8.0	72	844	4.6	26.0	1,110
	0.75	6.85	14.6	319	685	8.0	99	925	4.6	35.0	1,220
	1.0	7.37	14.6	401	710	8.0	124	960	4.6	43	1,270
6	0.0625	4.42	9.7	37.5	675	5.3	13.6	913	3.1	5.9	1,200
	0.5	6.82	9.7	159	840	5.3	50	1,130	3.1	18.0	1,490
	0.75	7.70	9.7	224	890	5.3	70	1,200	3.1	24.5	1,580
	1.0	9.30	9.7	284	980	5.3	88	1,320	3.1	30.5	1,740
	1.5	10.6	9.7	396	1,045	5.3	122	1,410	3.1	41.7	1,860
8	0.0625	5.41	7.3	22.9	860	4.0	8.7	1,160	2.3	3.9	1,530
	0.5	8.20	7.3	121	1,060	4.0	38.2	1,430	2.3	13.7	1,880
	0.75	9.59	7.3	172	1,145	4.0	54	1,550	2.3	18.8	2,040
	1.0	10.7	7.3	221	1,210	4.0	68	1,630	2.3	23.7	2,150
	1.5	13.1	7.3	311	1,340	4.0	95	1,810	2.3	32.7	2,380
	2	13.5	7.3	393	1,360	4.0	120	1,840	2.3	40.9	2,420

Note: All values apply if the thickness T is at least equal to $^1/_8$ inch. Resistance and reactance values apply and the remaining values very nearly apply if the thickness T is at least 0.035 inch for 10,000 cycles, 0.064 inch for 3,000 cycles and 0.111 inch for 1,000 cycles.



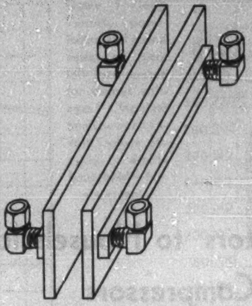


Figure 6 (left). Current carrying capacity per unit width for 1/4-inch thick copper bus bars on 10,000 cycles

A—2-inch wide bus bars
B—4-inch wide bus bars

Figure 7 (above). Water-cooled bus bars

this being the case of principal interest. Figure 5 gives a comparison between the experimentally determined current carrying capacity of 2-inch wide, 1/4-inch thick copper bus bars with *W* horizontal and *W* vertical, the frequency being 10,000 cycles. The current carrying capacity is greatest with *W* vertical, since in this position of the bus bars the air in the space between them is allowed to rise when heated, thus increasing the absorption of heat by convection. This effect is most pronounced if the spacing is equal to one-fourth the width, where the current carrying capacity with *W* vertical is about 15 per cent greater than with *W* horizontal.

Figure 6 shows current carrying capacity per-unit width plotted against the ratio of spacing to width for 2- and 4-inch wide bus bars. It is seen that if the width and spacing are both doubled, the current carrying capacity is approximately double also.

For a given transmission line which may be operated at two different frequencies, the product of the allowable current squared multiplied by the resistance per unit length is the same for both frequencies. This fact was used in Table II for extending experimentally determined current capacities for a frequency of 10,000 cycles to other frequencies. The current carrying capacities given may also be compared with those for 60 cycles, on the basis that the watts per inch of run for a 50 degrees centigrade rise should be the same for all frequencies.

Current carrying capacity values in Table II are for copper bus bars carrying a current in the horizontal direction with dimension *W* vertical. The color of the surface is taken as that of oxidized copper. If the surface is coated with black oxide of lead, the current carrying capacities are increased due to increased radiation. If the spacing is one-fourth the width, the allowable current is 12 per cent greater with a black oxide of lead surface than with an oxidized copper surface.

When a large current must be carried within a limited space, it may be necessary to use water-cooled copper bus bars. A typical construction is shown in Figure 7.

Conclusions

1. Equations have been presented for the calculation of resistance and reactance per-unit length of rectangular bus bar transmission lines. These equations have been checked experimentally for a frequency of 10,000 cycles.

2. Values of current carrying capacity have been determined experimentally for certain combinations of width and spacing at 10,000 cycles and extended to other combinations of width and spacing and to other frequencies.

3. For convenience, the results of 1 and 2 have been tabulated for frequencies of 10,000, 3,000, and 1,000 cycles.

4. If the thickness T is greater the 1.2 d where d is the reference depth, the resistance per-unit length is not appreciably decreased.

5. Maximum current carrying capacity with a given size bus results when the width W is vertical, the spacing S is at least equal to $0.25W$, and a coating of black oxide of lead is applied to increase radiation. If additional current carrying capacity is required, a larger bus or water cooling is required. There is no advantage to a spacing S greater than $0.25W$ with W vertical except where reactance is required for matching or there is danger of a voltage breakdown. Unless a strip of insulation is provided between the bus bars, the spacing should be not less than 1/2 inch for an rms potential difference of 440 volts or 220 volts and 3/4 inch for an rms potential difference of 800 volts, to avoid possibility of a voltage breakdown.

References

1. FORMULAS AND TABLES FOR CALCULATION OF MUTUAL AND SELF INDUCTANCE, E. B. Rosa, F. W. Grover. *Scientific Paper Number 169*, United States Bureau of Standards (Washington, D. C.), 1916, pages 156–57 and 169.

2. ANACONDA COPPER BUS CONDUCTORS. *Publication C-25*, American Brass Company (Waterbury, Conn.), fourth edition, pages 4–5, 8–10, and 16.

3. STANDARDS FOR INDUSTRIAL CONTROL. *Publications 45–97*, National Electrical Manufacturers Association (New York, N. Y.), September 1945, page 12.

4. THE DESIGN OF BUS-BAR INDUSTRIAL DISTRIBUTION SYSTEMS: AN EPITOMIZATION OF AVAILABLE DATA, T. J. Higgins. *AIEE Transactions*, volume 64, 1945, pages 384–400; discussion 486–88.

No Discussion

Application of Motors to Household Refrigeration Compressors

L. C. PACKER
MEMBER AIEE

HOUSEHOLD refrigeration has come a long way since the days when the ice came from a commercial ice-making plant, lake or river and was placed in an ice box. The first mechanical compressor consisted of a pump, belt driven by an electric motor. There were shaft seals in the pump to prevent the refrigerant and oil from escaping. The belt and shaft seals were troublesome. The manufacturers did the best they could to make as good a box, and the mechanism as quiet, efficient, and free of vibration as was possible at that time. However, with all its shortcomings, the housewife welcomed this type of refrigeration in preference to the old ice box.

One major step in solving these problems was the bringing out of the hermetically sealed unit into which was assembled the pump, motor, refrigerant, and oil. This design made it possible to eliminate belt, shaft, and seal, and thus provide more space in the box for food per unit area of floor space. Less maintenance is required because of elimination of the belt and shaft seal and by sealing all of the working parts. The complete unit is also quieter, has less vibration, and can be manufactured at a lower cost than the belt-driven unit.

The hermetically sealed unit is a mechanical, electrical, and chemical unit which poses many problems, and which has taxed the ingenuity of mechanical, electrical, chemical, metallurgical, welding, manufacturing, and tool engineers. Today, as a result of their co-operation, the industry can point with pride to the great progress made in the art of making hermetically sealed refrigerators for the household, as well as for commercial use. Professional artists have an important part in the styling of the cabinets.

When the housewife looks at the refrigerator in the store, she sees only the beauty of the appliance and its conveniences. She cannot see its performance, and while we all try to exploit it in selling, it is difficult to get her to understand it. However, a manufacturer who expects to stay in business must build in good performance, which means satisfactory temperature throughout the various parts of the box at minimum operating cost. The housewife expects and gets reliable performance year after year.

The trend is for larger refrigerators on the same floor space. This means that the larger food space and frozen food compartments will require better insulation and more efficient refrigeration units. Less space also is available for the motor and compressor.

It is not within the scope of this paper to discuss the complete motor design for household refrigerators, but it is necessary in making a successful motor application for the motor engineers to know:

1. The type of motor best suited for the application.
2. The power requirements and performance under all load and voltage conditions.
3. How the motor is to be started.
4. The maximum permissible motor temperatures.
5. That the motor is properly protected from overheating.
6. The chemical reactions that take place with various materials in the complete sealed unit.
7. The mechanical design of the motor and the materials necessary to satisfy the application.
8. The importance of mechanical, as well as chemical cleanliness.
9. The compressor design insofar as it affects the assembly of the motor.
10. The possible causes of motor failures in service which will not only be of considerable value to the motor designer, but also to the designers of the compressor and the other electrical equipment.

Motor Types

Neither d-c nor repulsion-type motors are applicable to hermetically sealed compressors. Squirrel cage rotor type a-c motors can be designed and built to handle almost any range required, except for exceptionally wide ranges of voltage.

Most household refrigeration systems are of the capillary tube type. Once a unit has stopped, it takes 3 to 5 minutes for the suction and discharge pressures to equalize, thus requiring a relatively low starting torque to start the compressor and get it up to speed and keeping the starting current within permissible operating limits.

The motors for household refrigerators are usually of the single-phase 4-pole 60-cycle 1,725-RPM type with split phase starting.

There may be cases where it is desirable to use a capacitor start motor to obtain more starting torque, but capacitor start motors are seldom used in household service.

The household refrigerator motor ratings vary from one-tenth to one-quarter horsepower. However, the horsepower rating does not mean much. The motor is designed for specific compressor applications.

Practically all of the 50-cycle circuits in this country have been changed over to 60 cycles. There is also very little 25-cycle current in this country. Domestic motors need only to be rated at 115 volts, 60 cycles. The export trade requires 115-,

---

Paper 52-69, recommended by the AIEE Domestic and Commercial Applications Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted October 23, 1951; made available for printing December 13, 1951.

L. C. PACKER is with the Westinghouse Electric Corporation, Springfield, Mass.

The author wishes to acknowledge the assistance and helpful criticisms of his associates in the Electric Appliance Engineering Department.

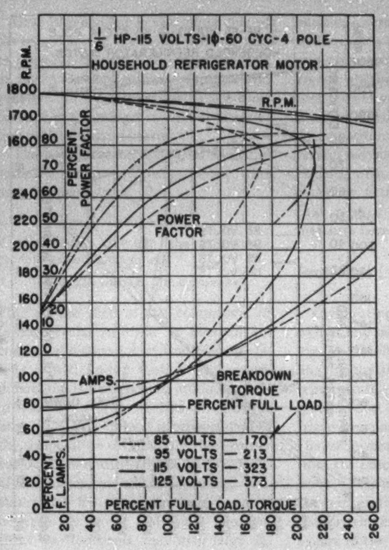


Figure 1 (left). Current, power factor, and speed versus per cent full load torque for various voltages. One-sixth horsepower 115-volt 1-phase 60-cycle 4-pole household refrigerator motor

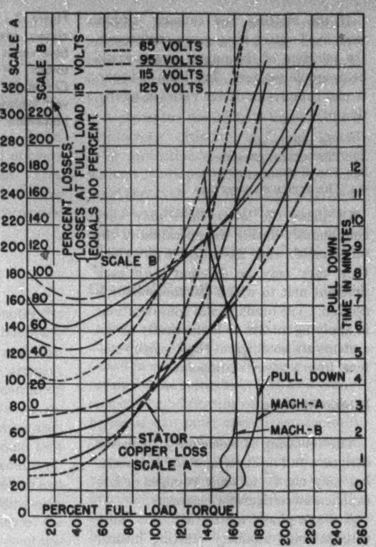


Figure 2 (right). Total losses, copper losses, and pulldown time versus per cent full load torque for various voltages. One-sixth horsepower 115-volt 1-phase 60-cycle 4-pole household refrigerator motor

200- to 250-volt 50- and 60-cycle motors. The most economical motor design for domestic use may not be satisfactory for 50-cycle operation. Reactors and transformers are available to use with standard 115-volt 60-cycle domestic units to operate on most of the voltages and frequencies encountered in export trade.

Power Requirements

The motor that drives the compressor must be considered as an important part of the refrigerator system and not an accessory. It must be adequate to drive the compressor to obtain the required amount of refrigeration.

The motor must have sufficient starting and accelerating torque to start and accelerate the load at the lowest expected voltage at the terminals.

The motor must have sufficient breakdown torque to handle the maximum load at the lowest voltage.

The motor must also operate at a satisfactory temperature under operating conditions for the required life and at an efficiency to give low operating cost consistent with the initial investment. It must do this under both minimum and maximum voltage conditions.

The difference between minimum and maximum voltages that may be encountered is too great a range for the successful operation of a refrigerator motor. Voltages as low as 70 have been found in some districts, and from there on up to close to 130 volts. If a motor were designed to operate satisfactorily from 70 to 130 volts, the motor would be excessively large, leading to excessive cost for the motor and compressor. Therefore, it is more economical and more satisfactory from a performance standpoint, if the motor is designed to operate on an average range from 100 to 125 volts, 60 cycles.

Curve Figure 1 shows what may be expected in variations in speed, current, and power factor, at different loads and at a large range of voltages. This particular motor shows only a 4 per cent spread in current at 100 per cent torque load on a range of 85 to 125 volts, but at lighter loads the current is higher at voltages above 115 volts and lower at voltages below 115 volts. The 85-volt curve is beyond the practical operating range.

At loads above 100 per cent, the current at voltages less than 115 increases at a much more rapid rate with decrease in voltage. In this particular design, the current at 125 volts is slightly above the value of current for 115 volts until it reaches about 130 per cent and then the current is less. Since the temperature rise of the winding varies approximately with the square of the current, it can be seen that the temperature rise increases very rapidly with increase in load at low voltages.

The breakdown, starting and accelerating torques vary approximately as the square of the voltage which means that at 85 volts, these torques will be approximately half of those at 115 volts.

Figure 2 shows the total losses, stator copper loss, and pulldown curves, plotted against per cent full load torque. The loss curves show how the losses vary with load and voltage. The stator copper loss reflects the current variation as shown in Figure 1. The total losses include the stator copper loss, rotor cage loss, iron and friction losses. The iron loss decreases as the voltage decreases but not in the same proportion.

The pulldown curves in Figures 2 and 3 show the torque versus time in pulling down the cabinet temperature to approximately the normal operating temperature. The torque will continue to drop slightly as the desired cabinet temperature had not yet been reached. Some cabinets require less and some require more torque for the pulldown than the two illustrations shown.

Figure 3 shows the watts input, efficiency, speed, and also pulldown time plotted against per cent full load torque.

The efficiency increases with decrease in voltage below 110 per cent full load torque and decreases with decrease in voltage above 110 per cent load.

The watts input decreases with decrease in voltage with loads below 110 to 120 per cent load and increases with increase in load above the 110 to 120 per cent load.

If the starting and running torques were fixed values for a given compressor, the selection of the proper motor would be a simple matter. The load of a particular compressor design varies as a function of:

1. The refrigerant.
2. The suction pressure.
3. The discharge pressure.
4. The degree of superheat.

Of these, the latter three may vary continuously during the operation of the refrigerator, and the range over which they vary may be different with each different unit to which the compressor is applied. The fundamental considerations in applying motors to hermetically sealed systems are no different than to open systems from a power standpoint.

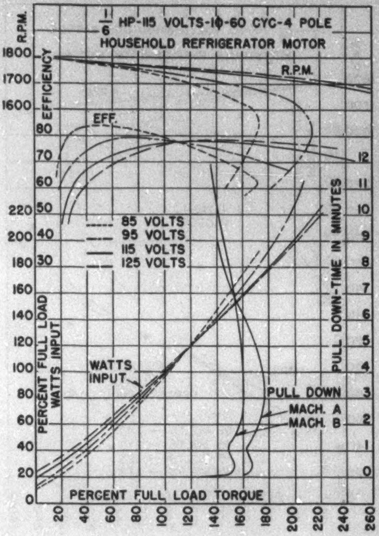


Figure 3. Watts input, efficiency, speed, and pulldown time versus per cent full load torque for various voltages. One-sixth horsepower 115-volt 1-phase 60-cycle 4-pole household refrigerator motor

Motor Starting

The following is elementary to motor designers and most refrigerator engineers, but may not be to some younger or less experienced refrigerator engineers.

A single-phase motor will not start without some auxiliary starting means. From a practical standpoint, it is necessary to wind into the stator core a starting winding displaced 90 electrical degrees with respect to the main or running winding. The combination of both windings produces sufficient torque to start and bring the load up to predetermined speed. The starting winding is then cut out and the main winding takes the load on up to speed. The starting winding is usually cut out at a speed just above the breakdown torque point of the main winding. If the load should become excessive, arrangements can be made to pick up the starting winding at a point below the breakdown torque of the main winding and take the load up to speed again because the torque of the combined main and starting winding is greater than that of the main winding only. Operation of this type of motor for short periods under heavy pulldown conditions at acceptable minimum voltages is permissible. However, too frequent operation under these conditions is detrimental to the life of the motor, contacts of the relay, and the protective device.

It is not permissible to have any transfer switching means such as centrifugal switches inside of the hermetic unit because of the inherent sparking, therefore, relays of various types must be mounted outside of the compressor for closing and breaking the starting winding circuit. These relays may be of:

1. The current type.
2. The time delay or hot wire type.
3. The voltage type.

Current Type Relay

This type is used in many household refrigerator applications. The current coil is placed in series with the main winding. It is designed so that after the line switch is closed, it will close the contacts at a voltage below the minimum voltage on which the motor is expected to start and run. The motor starts and accelerates the load and at a predetermined speed, the armature of the relay drops out, opening the contacts; thus disconnecting the starting winding. If the load becomes excessive and the speed drops below the breakdown torque of the main winding, a point is reached where the current increases to the closing current value of the relay. The relay armature picks up, closes the contacts, and connects the starting winding into the circuit to take the load up to speed again. This type of relay has been very successful.

Time Delay Relays

Time delay relays of the thermal type are also used in many household refrigerator applications for opening the starting winding circuit. One very common such type is known as the hot wire relay. The timing is such that it will open the starting winding circuit under the worst condition of load and voltage which may occur. Normally, the time is longer than necessary to get the motor up to speed. Although this may cause some noise in coming up to speed, this type of relay, too, has been very successful.

Voltage Type Relay

The voltage type relay normally has the contacts closed. The motor design must be such that the voltage across the auxiliary or starting winding will increase appreciably with speed. The increased voltage moves the relay armature to open the contacts. This type relay can only be used with capacitor start motors. The voltage change in the starting winding of a split-phase motor is inherently not sufficient to successfully operate this type of relay, therefore it is not used for household refrigerator motors.

In order to have successful operation of the electrical system, it is necessary to carefully select the relays to meet the characteristics of the particular motor involved. The complete electrical system must be very carefully co-ordinated for the best results.

Motor Temperature

The temperature of the motor and compressor parts are closely related because of their proximity. The presence of circulating oil and a relatively dense gas adds to the problem of temperature exchange between these closely related parts. This is particularly true with the small sizes of compressors and motors.

Heat is generated in the motor due to the copper loss of the motor winding, and to the eddy current and hysteresis losses in the motor laminations. Heat is also generated in the compressor by the compression of the refrigerant and by the friction in the common bearing.

The heat must be dissipated from the complete motor-compressor unit; this may be done by:

1. Radiation and convection.
2. Conduction through the compressor mountings and connections.
3. Flow of gas through the compressor.

Fins on the shell of the compressor or forced circulation of air by fans will increase the dissipation of heat from them.

It is well known that a temperature differential is necessary to produce a flow of heat to the point of dissipation of the heat. The flow of heat is proportional to the temperature differential, which causes various temperatures in the various parts of the motor, pump, and housing. Instances are known where the differential between the motor and pump temperatures was too great, and with large decreases in motor loss, its temperature decreased very little because of the heat flow from the pump into the motor. Therefore, it would be ideal in the construction of the complete compressor unit to have a heat barrier between the motor and the pump so as to have poor conduction between them, but leave good conduction of heat from each to the outer shell. Circulating the oil properly in the housing will help to transfer the heat to the outer shell. Providing means for circulating the cool incoming gas through and over the motor will greatly assist in carrying away the heat due to motor losses.

Moisture in the sealed refrigeration system is harmful in numerous ways, causing both electrical and mechanical troubles. All organic materials slowly deteriorate at normal temperatures, giving off water. This decomposition increases in rate with an increase in temperature. The products formed by the decomposition of cellulose at excessive temperatures consist of approximately 35 per cent water, 39 per cent carbon, and the balance in gaseous and scattered products.

In a hermetic system, this decomposition seems to start slightly above 200 degrees Fahrenheit, and increases at a somewhat higher rate than in air at higher temperatures. Overheating, even for short periods, should be avoided on hermetic motors even more than open motors where the products can escape to the atmosphere. Because overheating causes the liberation of moisture, it is difficult to examine a system which has failed and determine with any degree of certainty the original cause of the failure. It is not known whether there was excessive moisture present when the unit was manufactured or whether it was produced by the decomposition of the insulation. Even certain mechanical failures result in overheating, with the same end result. This makes it extremely difficult to analyze field returns and determine corrective measures necessary.

Protection from Overheating

Overheating of a hermetic motor, as well as any motor, is caused by overloading, undervoltage, overvoltage, and stalling of the motor. Abnormal ambient temperature, failure of the ventilating system, and in some cases, loss of refrigerant, contribute to overheating also. Hermetic motors also receive some heat from the pump. The motor does not receive a constant amount of cooling and the temperature rise of the winding over the air temperature does not vary as the square of the current. In some compressor designs, the temperature rise decreases with increase in load due to gas cooling.

The inherent overheating type of protection protects the motor against various causes of overheating previously mentioned. In this type of protection, as applied to hermetic motors, it is assumed that the temperature differential between the motor winding and the protector location varies approximately as the square of the current. This temperature differential is relatively small when the motor is running.

An inherent overheating protective device consists fundamentally of a heating element mounted near a temperature-responsive element. The assembled unit is mounted on the compressor-motor shell so that the temperature-responsive element responds to a combination of the heat from the compressor shell and the line current through the heating element. This gives protection during running conditions. During stalled or extremely heavy loads the heavy current through the heating element causes the temperature responsive element to open the line circuit to the motor even though the compressor shell temperature is low. The present types of inherent protective devices in use fully protect the starting windings against all conditions except welded relay contacts. The starting winding is wound with high resistance wire and its current density is so high that the temperature rises many times faster than the main winding. One of the present remotely mounted types of relays, the hot wire type, will also protect the starting winding under all conditions except welded relay contacts.

A protector should be selected to give satisfactory ultimate trip and locked rotor protection to the running winding without nuisance trip-out during the pulldown cycle and temporary overloads such as ice *freezing*. At the same time it must protect the starting winding under all running and stalled conditions.

Chemical Reactions

As previously stated, the motor and all compressor parts are completely enclosed in a shell which also serves as a reservoir for the refrigerant and as a crankcase for the lubricating oil. The complete hermetic unit must operate satisfactorily for a long period of time without attention. Therefore, the materials in the various parts must not undergo any appreciable change during this time. This includes all of the metals involved, lubricating oil, and all motor insulation.

Most household refrigerators now use Freon-12 or Freon-114 refrigerant, but whether it is sulphur dioxide, methylchloride or Freon-12, there are corrosive effects which have been found by laboratory tests to occur in both wet and dry refrigeration systems. However, the chemical reaction is much greater in wet systems and more so in the methyl chloride system. During World War II, there was very little Freon-12 available for refrigeration systems and should a similar occasion arise in the future, it may be necessary to use another less critical refrigerant if refrigerators are to be made. Therefore, the following information based on 24 hours exposure at 125 degrees centigrade may be of some interest to the motor and refrigerator manufacturers.

METHYL CHLORIDE, DRY SYSTEMS

There is no chemical reaction with steel, copper, brass, and phosphor bronze. Aluminum decomposes. Cotton chars and paper insulation chars and becomes brittle. Oil darkens if cotton is immersed in it.

METHYL CHLORIDE, WET SYSTEMS

Steel is corroded black and collects copper plating. Aluminum corrodes and becomes a fire and explosive hazard. Copper, brass, and phosphor bronze corrode. Cotton and paper insulation chars. Paper insulation and enamel on wire become brittle. Oil darkens in the presence of cotton. The chemical reaction on cotton and paper in dry methyl chloride is greater than the other systems and the addition of a small amount of water causes the corrosive action to be particularly bad. It has been found that it takes a smaller amount of water in methyl chloride systems to accelerate chemical reaction than in other systems.

SULPHUR DIOXIDE, DRY SYSTEMS

There is no reaction with steel, aluminum, copper, brass, phosphor bronze or paper insulation. Cotton insulation turns brown. Enamel wire is attacked and the oil darkens.

SULPHUR DIOXIDE, WET SYSTEMS

Steel corrodes black. Copper, brass, and phosphor bronze corrode. Cotton insulation blackens and paper insulation deteriorates and becomes weaker. Enamel on the wire is attacked and the oil darkens.

FREON-12, DRY SYSTEMS

There is no reaction with any of the materials mentioned.

FREON-12, WET SYSTEMS

There are traces of corrosion on steel, and a white discoloration on aluminum. Copper and brass have a black discoloration. No reaction with phosphor bronze. Cotton and paper insulation show a slight deterioration. Enamel wire and oil are not attacked.

Magnesium must not be used in any of these refrigerants.

It is imperative that all materials be as clean chemically as possible, and that there be no extractable materials present that will react to cause corrosion or deterioration.

A very small amount of moisture may be present because it is not possible to obtain a completely anhydrous condition in a refrigeration system. However, both assembly and dehydrating practice must be closely controlled to keep the moisture within limits which will not cause the system to freeze up under any operating condition.

Fractional horsepower motors for household and commercial refrigeration are manufactured with the stator windings unimpregnated because the varnish slows up the rate of dehydration. Oil should also be kept off the windings during the dehydration cycle. In case a unit is reworked, the oil should all be removed by washing in a solvent.

Once all motor materials, including insulation, have been established as suitable for a refrigeration system, no material changes should be made without first testing under actual operating conditions for a period sufficiently long to prove that the new materials are satisfactory for such use.

Mechanical Design of Motor Parts

The motor manufacturer usually supplies only the complete wound stator and rotor, less shaft.

The motor industry through the National Electrical Manufacturers Association has standardized on the stator outside diameter with tolerances, stator widths over the winding, and rotor bore with tolerances which permit the compressor manufacturer to establish standard dimensions and tolerance for the proper shaft fits in the rotor and the fit between the stator core and housings. This latter fit plays an important part in the transfer of the heat through the housings. The fit must be tight enough to hold the stator in place in the housing, even in vertical units, yet it must not be so tight that it will distort the stator.

There must be sufficient rigidity in the stator construction to withstand ordinary handling without disturbing the close machining tolerances, thus facilitating assembly by the compressor manufacturer and promoting quality apparatus. Either welded or riveted stator cores or a frame wrapped around the stator core may be used, depending upon the features desired in the individual designs.

As the winding in the stators for household refrigerator compressors are not impregnated with varnish, they will loosen at the end winding and also shift slightly in the slots. They are tied, but it is impossible with the type of insulation used to tie down the wires so that there is no relative motion in service after the paper and cotton are saturated with oil. Tightly wedging the wires in the slots minimizes movement in the slots.

The insulation must be so designed that no abrasion will result and cause short circuits or grounds. The motor manufacturer should maintain outline dimensions of the stator winding as consistent as possible with the nature of the winding.

The motor laminations are made from the lowest core loss, highest permeability, and greatest thickness electrical sheet steel consistent with the high performance necessary, low cost and good punchability of the steel. The refrigerator manufacturer is interested in both high efficiency and low costs. The steel must be free of any covering that will contaminate the refrigerator system, and be free of oil. In most cases, the stator laminations are annealed to improve performance.

The stator windings may be put in by hand, but in many cases the main winding at least is wound by machine, which results in greater accuracy in the number of turns and lower cost. In practically all cases, the main and starting windings are insulated with enamel. This has resulted in more efficient designs, compared with previous designs using double cotton covered wire.

In many cases, the completed stator is washed to cleanse it of foreign matter. The stator is then tested with equipment that detects short and open circuits in the main and starting winding and between windings, as well as detecting the wrong number of turns and grounds to the stator core.

The rotor is manufactured of the same grade of steel as the stator core and with either copper or die cast aluminum cage. It should be noted that aluminum free of magnesium and impurities is suitable for use in Freon-12 or sulphur dioxide, but not in methyl chloride systems. The laminations are annealed before being assembled into cores for die cast rotors, and in many cases for copper cage rotors. This eliminates the stray load losses, which cause dips in the speed torque curve, and which reduce the pull-up and breakdown torques and efficiency. After the rotor is cast and the rotor surface and bore are finish machined, the complete rotor is then given a special heat treatment to reduce the stray load losses caused by die casting and machining operations. Many copper cage rotors also get this treatment.

Test equipment has been devised to detect stray load losses, to check the speed torque curve, to detect high resistance and broken bars, to indicate the wrong skew of the slots, and to show excessive eccentricity of the rotor. All of this equipment can be arranged to give a visual indication. Some motor manufacturers give the motor a dynamometer test before shipping it to the compressor manufacturer.

After test, the stator and rotor are wrapped in special cellophane bags to keep out moisture and foreign matter and then packed for shipment.

Cleanliness

One of the most important items necessary in manufacturing hermetically sealed units is cleanliness. The motor and compressor manufacturers have separate, as well as joint, responsibilities in maintaining cleanliness of the motor and compressor parts and the complete assembled unit.

The motor must be free of dirt, rust, paint, grinding dust, metal chips, and other insoluble matter as they will clog up small refrigerator passages, valves, and screens as well as damage the bearing surfaces and valves. No dirt should be permitted to collect between the laminations or in the winding as it may come out in service and get into the refrigerator system.

No material must be used in the motor which can be attacked by oil, or oil vapor and the refrigerant; material must not contain any soluble impurities.

To date, the only practical insulation for the motor has been of the organic or AIEE class *A* type. Paper is used for slot cells, barriers, slot wedges, separators, connection and cross-over wire insulators.

Cotton is used for sleeving over lead wires, and for binding the winding. The stator wire is usually covered with a polyvinyl acetal or other types of enamel. Sometimes the enamel has a cotton covering. These materials in the refrigerator system require extreme mechanical and chemical cleanliness, and a high degree of dehydration. For instance, all paper and cotton products must be free of waxes and other impurities which may be dissolved and deposited in the small passageways. They must also be free of lint, as it will clog screens in the refrigerant line. Wire enamels and such materials must be selected for freedom from the effects of oil and the refrigerant. This includes impregnating varnishes although they are seldom, if ever, used in household refrigerator motors. Core plate varnish or an excessive amount of oil on the motor laminations is not permissible.

There is also a joint responsibility in the selection of materials used in the motor and the balance of the refrigerator system so that no combination of the basic materials or their impurities will result in corrosion or products that will produce objectionable effects in the system.

It is the compressor manufacturer's responsibility to make certain that the compressor and other parts of the system are free of materials that will damage the motor. For instance, metal chips can readily short circuit the winding. Cast iron and steel chips are easily attracted by the magnetic field and get into the winding. The compressor manufacturer should store the motor parts, prior to using them, in clean storage rooms of low humidity to prevent rusting. He should also handle them during assembly so as to avoid finger stains on steel parts, or rust will soon appear. He must avoid getting dirt or oil in the winding because it may not be readily removed and may later get into the refrigerator system. He must handle the stators carefully so that the winding does not become damaged.

Compressor Construction

As was previously stated, the hermetically sealed compressor includes the pump oil, refrigerant, and an enclosing shell. It is a more simple construction than the belt driven type. Common bearings are used for both the pump and motor and they are designed to provide adequate capacity for the combined load. The crankshaft of the pump is integral with the motor shaft. An adequate lubrication system must be supplied with sufficient oil to last the life of the unit.

Proper clearances must be maintained between the bearings and shaft. If clearances are too close, the shaft will freeze in the bearing under extreme temperature or load conditions. If too liberal, the shaft will rattle in the bearings. It is extremely important to maintain proper tolerances and surfaces in the bearings and on the shaft, otherwise the result will be more rejections in the factory and high service charges.

The unbalanced magnetic pull and noise from the motor increases rapidly with any uneveness of the air gap between the motor stator bore and rotor. This pull increases the pressure on the bearings and may break down the oil film. It also increases the stresses in the shaft. Therefore, the construction of the housing must be such as to permit accurate machining and assembly.

An excess of oil in the air gap of the motor causes a drag on the rotor and requires more power input to maintain a given load.

The compressor manufacturer must provide adequate clearance between the stator winding extensions and the housing, recognizing the fact that the winding extensions cannot be formed to maintain as definite a shape as a solid section.

Motor Failures in Service

The following motor failures may occur in service:

1. The main winding may fail due to various causes

   Faulty wire that may pass the most meticulous shop test but show up through vibration or injure the enamel because of contraction and expansion caused by temperature change when put into service.

   Metal particles in the winding.

   Nicks and abrasions on the winding and its insulation caused by rough handling.

   Test equipment which may not detect marginal cases which may later show up in service.

   Improper manufacture of motor and compressor parts, assembly, and soldering of leads.

   Excessive burrs in or at the ends of the stator slots may, in time, break through the insulation due to pressure and vibration.

   Failure of the thermostat.

   Excessive amount of water in the unit caused by improper dehydration or by overheating of the motor.

   Improper control of the dehydration ovens causing overheating of the motor insulation.

   Abnormal voltages.

   The ultimate results are short circuits, open circuits, and grounds.

2. The starting winding may fail due to the failure of the relay or thermostat and the reasons given for main winding failures. If a failure is caused by excessive heating, it may affect the main winding and cause it to fail. Under ordinary conditions of starting, the starting winding will not become heated excessively, but since the design characteristics are such that the current density in this winding is very high, defective thermostats or relays that cause the current to flow for a longer time than the normal starting cycle will result in damage to this winding. Repeated starts with insufficient time for equalization of the pressures may also damage the winding.

3. Failures caused by improper processing of die cast rotors may be due to

   Excessive stray load losses resulting in a cusp in the speed torque curve and difficulty in getting the load up to speed. This causes excessive time of the current in the starting windings, especially at low voltages.

   Broken and/or high resistance bars which affect performance.

4. Sticking shafts may be caused by

   Lack of lubrication for one cause or another.

   Breakdown of the oil caused by chemical reactions in the system.

   Roughening of journals or bearings from corrosion.

5. Noise. The various types of objectionable noise caused by the motor may be due to these causes:

   Excessive bearing clearance.

   Poor finish of the shaft and bearing.

   Rotor not properly lined up with the magnetic center of the stator.

   Rotor not concentric with the bore of the bearing and stator.

   Broken bars and possibly high resistance bars in the rotor.

   Poor motor design. A motor which for any reason is designed too small for a given rating will not be as quiet because the higher magnetic densities will cause an increase in the magnetic noise. The motor may also be responsible for other parts of the system vibrating and causing objectionable noises. Any increase in magnetic densities will increase these noises.

Conclusions

The author has attempted to emphasize the requirements and problems involved in applying motors to household refrigeration compressors. This also includes the associated electrical equipment, the proper application of which is essential for a successful motor application. The success of the household refrigerator is evidence enough for the engineers and chemists responsible for the development and application of the motors, associated electrical equipment, and compressors to feel proud of their achievement. The same fine co-operation that was responsible for the progress to date will continue to show progress in the future. We may look forward to even higher quality, longer life, and lower cost designs. Higher motor efficiencies may be expected.

It is very desirable to have insulating materials that would not be affected by heat or by the combination of refrigerant and oil, or by any chemical reaction resulting from the presence of moisture, to replace the present organic types of insulation. Many new materials have been made which look promising, but it takes 2 to 3 years to prove them in actual service and to develop them into practical and economical materials.

References

1. Design Factors Involved in the Design of Motored Appliances, **L. C. Packer.** *AIEE Transactions*, volume 60, 1941, pages 500–07.

2. Electrical Equipment for Hermetically Sealed Refrigerators, **L. C. Packer, F. J. Johns, E. P. Codling.** *Refrigerating Engineering* (New York, N. Y.), June 1945.

3. Hermetic Electric Motors, **B. O. Haun, Jr.** *Refrigerating Engineering* (New York, N. Y.), August 1947, pages 135–38.

4. Motor Designs for Hermetic Units, **F. H. Fleischer, J. B. Brightman.** *Refrigerating Engineering* (New York, N. Y.), October 1951, pages 963.

5. Refrigerator Motors Need Thermal Protection, **E. B. Bremer, C. G. Veinott.** *Refrigeration News*, Detroit, Michigan, October 11, 1933.

6. Fundamental Theory of Inherent Overheating Protection Under Running Overload Conditions, **C. G. Veinott, L. C. Schaefer.** *AIEE Transactions*, volume 68, part I, pages 266–72.

7. A Method of Selecting Inherent-Overheat Protectors Based upon Motor Design Information, **V. G. Vaughan.** *AIEE Transactions*, volume 67, part II, 1948, pages 1023–30.

8. Drive Motor Torque Must Match Compressor Needs, **Henry B. Pownall.** *Refrigerating Engineering* (New York, N. Y.), May 1948, pages 457–51.

9. Compressor Design, **R. E. King.** *Refrigerating Engineering* (New York, N. Y.), May 1948, pages 461–63, 498.

10. Rating and Application of Motors for Refrigeration and Air Conditioning, **P. H. Rutherford.** *Electrical Engineering* (*AIEE Transactions*), volume 58, October 1939, pages 519–27.

11. The Inherent Overheating Protection of Single-Phase Motors, **C. P. Potter.** *Electrical Engineering* (*AIEE Transactions*), volume 60, November 1941, pages 993–96.

12. Overload Protection of Electric Motors, **E. F. Kurtz.** *Refrigerating Engineering* (New York, N. Y.), May 1944.

13. Fractional Horsepower Electric Motors (Book), **C. G. Veinott.** McGraw-Hill Book Company, New York, N. Y., 1948, 572 pages.

14. Refrigerating Data Book. American Society of Refrigerating Engineers, New York, N. Y., 1951.

Discussion

A. P. White (Metals and Controls Corporation, Attleboro, Mass.): Mr. Packer is to be commended for his excellent discussion of the various details which must be observed in the manufacture of hermetically sealed refrigeration compressors and the various problems which must be taken into account in providing a satisfactory design.

As the writer of these comments is associated with over-temperature protection of motors, this discussion will be limited to that part of Mr. Packer's technical paper dealing with over-heating protection.

Mr. Packer clearly points out that the motors in hermetically sealed refrigeration compressors unequivocally require overheating protection. The question raised though is the degree of protection obtained with existing types of protective devices.

A thermal overload relay mounted remotely from the hermetically sealed compressor operates in response to motor current only and its operating characteristic is given a curve proportional to the square of the current. Mr. Packer points out that hermetic motors do not receive a constant amount of cooling and, therefore, the temperature rise of the winding over the air temperature does not vary as the square of the current. This points out why thermal overloads located off of or away from the compressor shell are ineffective in protecting motor windings to definite temperature limits.

On the other hand, dome mounted thermal protectors are responsive not only to motor current but also to motor temperature as well by virtue of conduction of the heat of the winding out to the motor shell. Therefore, operation of the dome mounted or inherent overheating type of protector integrates both motor current and motor temperature in functioning to limit the motor winding to definite temperature levels.

One of the protection problems which Mr. Packer raised was operation with the starting switch contacts welded closed or failure of the start switch contacts to open because of low voltage conditions. Mr. Packer is correct in pointing out that the usual types of remote or inherent protective devices now used do not provide protection for the condition of the motor running with both the main and start windings energized. However, designs of inherent over-heating protectors can be furnished to provide protection for the condition of running on both windings as well as protection for conditions of running main winding only, stalled rotor main winding only, and stalled rotor both main and starting windings.

G. S. McCloy (Westinghouse Electric Corporation, Springfield, Mass.): In a discussion of the pulldown curves on the domestic refrigerator, it would be helpful to consider the mechanics of a pulldown on a typical domestic refrigerator. We will assume that we are dealing with a design that is generally in use today. It is a combination compressor-motor assembly pressed into a steel shell, a portion of which serves as the oil reservoir. The compressor intake is from the shell so the shell is normally maintained at substantially the same pressure as the low side of the compressor. The compressor discharges into a high side of relatively small volume and the liquid refrigerant is metered to the evaporator through a capillary tube type of restrictor. The total refrigerant charge is relatively small, being somewhat in the neighborhood of 16 ounces and is one of the Freon groups which is completely miscible with oil. Prior to the pulldown, stabilization is achieved by maintaining the refrigerator at a constant elevated room temperature for a sufficient period of time to obtain uniform temperature throughout. Under these conditions, the major portion of the refrigerant will be dissolved in the oil. That portion which is not dissolved will be in the vapor phase and occupying the volume of the system not occupied by the oil.

The major factor in the pulldown torque required is the pressure difference between the low and high side of the compressor.

At the instant of starting, there exists no pressure difference. Immediately after starting, a pressure difference is produced by the action of the compressor, in lowering the pressure in the shell, causing evolution of refrigerant from the oil, and discharging the compressed vapors into the high side condenser.

The discharge pressure rises rapidly until condensation begins. If the refrigerant is permitted to evolve freely from the oil, the pumping rate of the compressor can exceed the metering capacity of the capillary tube under normal conditions. This results in a backing up of the liquid refrigerant in the condenser, reducing the effective condensing area and thereby continuing to increase the discharge pressure. The suction pressure will fall as the refrigerant is pumped from the oil but there is introduced another factor to retard this rate of fall of the suction pressure. This latter factor is the effect produced when the refrigerant is discharged from the metering device into the warm evaporator which rapidly evaporates the entering refrigerant. These conditions persist until a maximum point is reached, requiring the maximum torque from the motor. The time during the pulldown period at which the maximum condition will occur will be determined by the rate of decline of the suction pressure and rate of increase of the discharge pressure. The major factors influencing the suction pressure decline will be the quantity of oil, the quantity of refrigerant, the ambient stabilized temperature, design of compressor, the type and volume of the low side, the rapidity of evolution of the refrigerant from the oil, the capillary tube or other restrictor design, and the volume of the high side. The rate at which the high side pressure will build up is a function of substantially the same factors just mentioned for the low side, plus the additional factor of the maintained ambient temperature.

It will be noted from the curve in Figure 2 and Figure 3 that the conditions requiring maximum torque during pulldown occurred approximately 3 minutes after the start.

L. C. Packer: The author appreciates the comments of Messrs. McCloy and White. They are valuable additions to the paper.

Mr. McCloy explains very clearly the causes for the torque requirements during the pulldown cycle of a refrigeration system. Motor engineers as a rule are not familiar with this cycle of operation, and Mr. McCloy's explanation should clear up this point.

Mr. White's comments extend the information in the paper and emphasizes the importance of the over-heat protection device and its application to properly protect the motor under all load and voltage conditions.

Increased Losses in a D-C Motor When Operated From Grid-Controlled Rectifiers

C. RICHARD REITER
ASSOCIATE MEMBER AIEE

CHARLES R. AMMERMAN
ASSOCIATE MEMBER AIEE

Synopsis: The increase in copper losses and iron losses of a d-c shunt motor when operated from grid-controlled rectifiers is studied. Fourier components of the rectifier output voltage are calculated, and the superposition principle is applied to predict copper losses. Tests were run to determine copper loss and iron loss; increased copper loss is seen to closely approximate the predicted value, and discrepancies are interpreted as being due to nonlinearity of the iron. The increased iron losses follow a pattern, explained in the text. Results are graphed on normalized co-ordinates and are thought to be representative.

WHEN d-c motors are operated from grid-controlled rectifiers, there is an increase of the losses in the motor which results in higher operating temperatures. The increased losses result from the ripple which is inherent in the output of grid-controlled rectifiers, and are localized in the electric and magnetic circuits of the machine.

A theoretical analysis of the voltage and current wave shapes supplied to a d-c motor when operated from a grid-controlled rectifier was conducted. An attempt was made to find the effect of various component wave shapes on the losses occurring in the motor during operation with the rectifier. Although the restrictions of the superposition theorem to linear circuits is well known, this theorem has been applied here in order to obtain an approximation to the actual currents. Since the machine has an air gap of appreciable magnitude and since the alternating components of current are usually considerably less than the steady or d-c components, it was thought that theoretical values would be comparable with measured values.

For the purpose of clarity in the analysis of the voltage and current wave shapes while supplying a d-c motor, the motor has been replaced by its equivalent load of a resistance, inductance, and battery representing the motor counter-electromotive force, in series, see Figure 1.

Consider first the voltage applied to the armature circuit Figure 2 compares the possible conditions. In Figure 2(A), there is no control of rectifier conduction; diodes produce this type of wave form. If a thyratron or ignitron is used, conduction through a particular phase may be delayed by making use of the grid or ignitor control. In an extreme case, as in 2(B), the delay is so great that current has stopped flowing part of the cycle. As the voltage due to one phase falls below the counter electromotive force (emf), conduction is maintained by the inductive voltage, $L\,di/dt$, developed in the armature. Eventually this will fail to be sufficient to maintain conduction and the terminal voltage jumps to the value of the counter-electromotive force as the current ceases to flow. The counter-electromotive force is practically constant as the machine has almost no time to slow down. At some time T_1 proper conditions are applied to the rectifier of the next phase and conduction starts. T_1 can be any time after the anode voltage of this particular rectifier rises above the counter-electromotive force. T_2 is the time at which this phase current ceases.

If T_1 is somewhat earlier, that is, if T_1 is before the previous diode automatically ceases, then the T_1 for one phase coincides with the T_2 for the previous phase. Current is always flowing through the load in this case, illustrated in Figure 2 (C). The earliest possible T_1 will occur in a case similar to that illustrated in Figure 2(A), where current is commutated between phases when the conducting phase voltage falls below the voltage of the succeeding phase. The time difference between T_1 for cases 2(B) or 2(C) and the earliest T_1 illustrated in 2(A) is called the delay time, and $\Delta w T_1 = \theta$ is the delay angle. The theory herein presented is general for a p-phase rectifier. Test and graphs apply to 3-phase (half-wave) rectification. It is not known whether the theory is born out for other values of p; but if p is greater than 3, the a-c components of voltage and current will be reduced. It would seem logical under these circumstances to assume that the nonlinearities would not be so significant in such a case, but on the other hand, the lowest a-c component frequency is $60p$ so that inductance and resistance values might be different than expected.

Theory

In this analysis, in which only continuous currents are considered, the wave shape of the voltage curve is represented by the formula

$$e = E_m \sin(wt + \theta) \tag{1}$$

where wt goes from 0 to $2\pi/p$ radius if p is the number of phases, half wave. This periodic wave shape is repeated every $2\pi/p$ radians and a voltage wave as shown in Figure 2(C) is formed.

Through a Fourier analysis the d-c component of voltage is found to be

$$E_{dc} = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} (E_m \sin[wt + \theta] - E_t)\,dt \tag{2}$$

Paper 52-74, recommended by the AIEE Industrial Control Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted October 23, 1951; made available for printing December 27, 1951.

C. RICHARD REITER is with the Shell Oil Company, Elk City, Okla., and CHARLES R. AMMERMAN is at The Pennsylvania State College, State College, Pa. At the time this work was done, Mr. Reiter was a graduate student in the Department of Electrical Engineering, Pennsylvania State College.

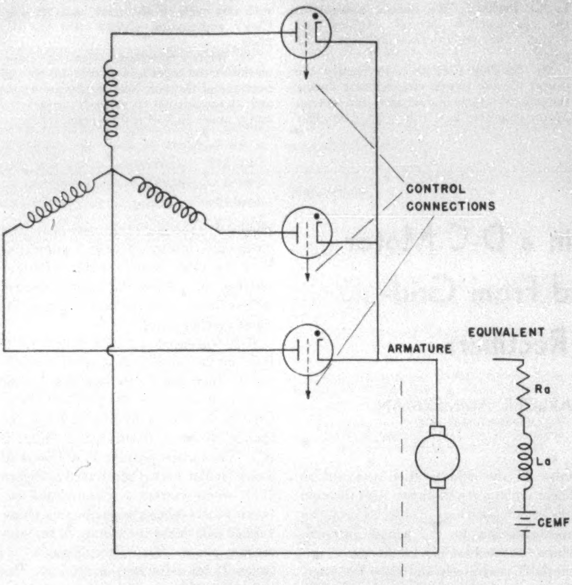


Figure 1. Circuit of rectifier and equivalent motor circuit

where E_t is the tube drop, which yields upon integration ($p=3$)

$$E_{dc}=\frac{3E_m}{2}\left(\cos\theta-\cos\left[\frac{2\pi}{3}+\theta\right]\right)-E_t \quad (3)$$

The sine wave components of voltage are found by solving

$$E_{\sin}=\frac{2}{T_2-T_1}\int_{T_1}^{T_2}E_m\left[\sin(wt+\theta)\times \cos\frac{2\pi nt}{T_2-T_1}\right]dt\,\sin\frac{2\pi nt}{T_2-T_1} \quad (4)$$

and the cosine wave components are found by solving

$$E_{\cos}=\frac{2}{T_2-T_1}\int_{T_1}^{T_2}E_m\left[\sin(wt+\theta)\times \sin\frac{2\pi nt}{T_2-T_1}\right]dt\,\cos\frac{2\pi nt}{T_2-T_1} \quad (5)$$

where n is the multiple of the fundamental a-c component being investigated, T_1 is the time of firing, and T_2 the time of firing next phase. The frequency of the fundamental component is equal to fp; here, 180 cycles per second. Since n can assume any value, there are an infinite number of components making up the voltage wave shape. In this specific case it was found that the equivalent current wave shapes were closely approximated when only the 180-cycle and 360-cycle frequency components ($n=1$ and $n=2$) were considered. The resultant equation for the voltage, using the 180-cycle and 360-cycle a-c components, is

$$\begin{aligned}e={}&E_m[0.4775\{\cos\theta-\cos(120^\circ-\theta)\}]-E_t\\&+E_m[-0.1035\sin\theta-0.1788\cos\theta]\cos\frac{2\pi t}{T}\\&+E_m[-0.0236\sin\theta-0.0136\cos\theta]\cos\frac{4\pi t}{T}\\&+E_m[0.5370\sin\theta-0.3100\cos\theta]\sin\frac{2\pi t}{T}\\&+E_m[0.2452\sin\theta-0.1418\cos\theta]\sin\frac{4\pi t}{T} \quad (6)\end{aligned}$$

where $T=1/180$ second.

Given the values of armature resistance, armature inductance (at a particular field current) and rectifier constants, the components of current in the circuit due to each of the various voltage components can be found. Applying the superposition theorem, it is a simple matter to calculate the copper loss which is equal to the sum of the I^2R losses of the individual components. It is apparent that the power useful in causing rotation is the d-c power. For a given speed and power developed therefore, the d-c component of current would remain fixed. The output would vary, however, as the rotational losses increase with alternating current. The extent of this increase is not easy to see, but it must certainly be less than the total core loss increase. For a given output, therefore, slightly greater direct current would be required using rectifiers than with constant voltage supply. This will cause additional I^2R loss but it will be a small second order effect which has been neglected here. The increase in IR loss is calculated by subtracting the $I_{dc}{}^2$ from the $I_{rms}{}^2$ and multiplying by the resistance.

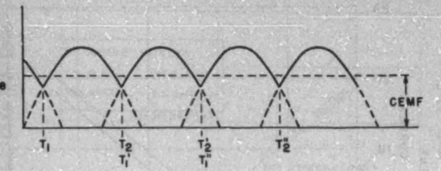
The calculation of iron loss for the simplest of cases is difficult and only approximate. In this case it is difficult to predict anything other than a trend. The actual increase for each motor must be obtained from tests. It is logical to assume that the loss would increase as the a-c components increase.

Test Procedure

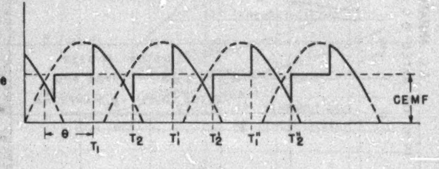
Standard AIEE tests were conducted on two machines to find the magnitude of normal friction losses, copper losses, core losses, and stray load losses for various speeds, loads, and flux conditions. To simplify analysis, it was desirable to use shunt motors in the tests. The machines used were designed for compound operation, but because of the inability to obtain shunt machines of suitable size for the tests, they were operated shunt by disconnecting the series field. The second machine was used as a generator loading the motor under test. The machines were rated 1$^1/_8$ kw as generators; based on the same armature current and same losses, rated output as a motor would be 1.3 horsepower. Normal field current was determined from a standard no load saturation curve, see Figure 3.

After standard tests using a d-c generator (constant voltage) power supply, the motor was operated under similar conditions with the power supplied by means of a Westinghouse grid-controlled rectifier. The difference between the losses when operated from the rectifier supply and when operated from the d-c generated supply represents the increased loss.

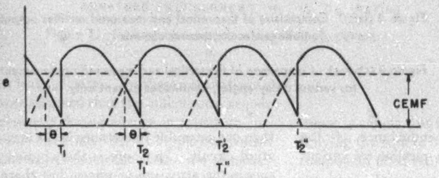
As previously stated, the tests during rectifier operation were limited to operation during periods of continuous current in the rectifier. With little or no delay continuous current was obtainable with a very light load. However, as the delay increased, it was found necessary to have larger loads in order to achieve the continuous currents. If conduction is delayed sufficiently, the current will be discontinuous regardless of the load. Ex-



t →
(A)
NO DELAY



t →
(B)
DISCONTINUOUS CURRENTS



t →
(C)
CONTINUOUS CURRENTS

Figure 2. Voltage wave forms

(A) No control
(B) Discontinuous currents
(C) Continuous current

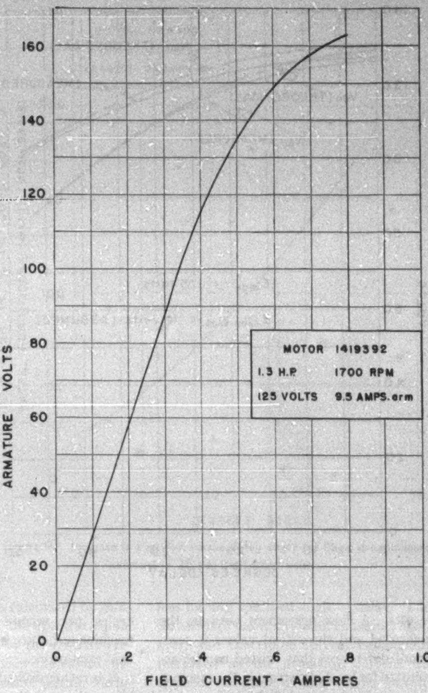


Figure 3. Saturation curve at rated speed of motor under test

perimental tests were conducted from no delay to a delay of 40 electrical degrees. It was discovered that field excitation was limited to approximately rated value because of occasional rectifier missfire. When the field current was reduced below approximately 80 per cent rated value, the motor operation became unstable. Each missfire caused speed variations of sizable magnitude in the motor and excessive currents. When the field current was increased to approximately 120 per cent rated value, the desired speed range could not be obtained. However, it was possible to find the effect on the armature inductance of field excitation variations from 0 to 200 per cent and to predict operation over this range.

Mechanically there was no perceptible difference in the operation of the motor when operated from the two different sources of power. The moment of inertia of the motor was apparently sufficiently large to smooth out any mechanical pulsations attributable to rectifier supply. It is, therefore, apparent that the friction loss remained constant for any given load or speed, regardless of the source of supply. Any additional loss that occurred when the motor was operated from the rectifier was either additional copper loss, core loss, or stray load loss. These losses are grouped as increased copper loss and increased iron loss.

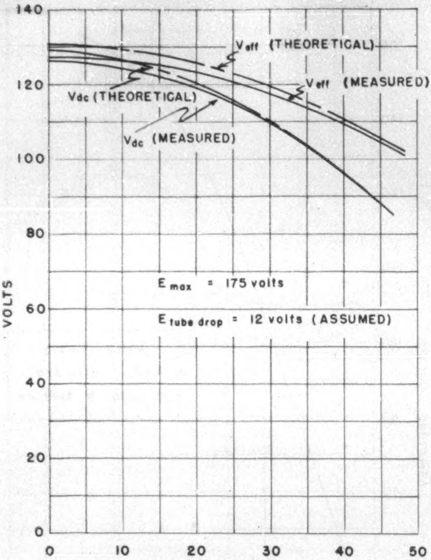
A dynamometer ammeter was used for measurement of the effective current. As explained, the increase in copper loss is taken as the difference between the I_{eff}^2R and the I_{dc}^2R. The remaining increased losses are allocated to the increase in iron loss, combining increased core losses and increased stray load losses.

The effect of supplying the field from the rectifier instead of constant voltage generator while the armature carried the rectified supply was checked. It was observed that there was no apparent change in the power consumed by the system over a complete range of loads. There was a ripple in the field current that was observed by means of an oscilloscope for both the d-c generated and rectified supply, but it was so small that it did not affect the ammeter readings. The ripple occurring when the field was supplied from a generated d-c source was set up by transformer action from the armature circuit.

Results

The rectifier transformer had a transformation ratio of 115/119.5 volts. The voltage supplied to the rectifier was approximately 119 volts rms and the rectifier E_m was 175 volts. The tube drop was assumed to be 12 volts.

Theoretical voltages obtained from the Fourier analysis presented are shown graphically in Figure 4, along with the test



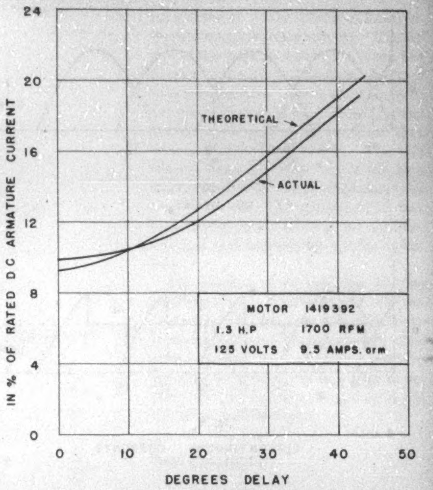


Figure 4 (left). Comparison of theoretical and measured rectifier output voltages for continuous currents

Figure 5 (above). Comparison of theoretical and measured ripple current for various delay angles, continuous current only

results. A close agreement between the measured and theoretical values is seen. Small differences that existed can be accounted for by slight errors in the assumption of tube drop and slight variations in the magnitude of the supply voltage. The small voltage drop in the line would also affect voltage readings obtained.

The d-c component of current is equal to the difference between the average voltage supplied the circuit and the back emf of the motor divided by the resistance of the armature circuit. To obtain the effective impedance for each voltage component, one must find the inductance of the armature circuit as well as the resistance. The total effective ripple current is then calculated by means of the following equation

$$I_{\text{ripple}} = \sqrt{I_{1ac}^2+I_{2ac}^2+I_{3ac}^2\ldots+I_{nac}^2} \quad (7)$$

Figure 5 graphically shows the theoretical and actual values of ripple current obtained. It can be seen that the two curves closely approximate each other.

Armature inductance was calculated after measurements of impedance and power loss were made on the stationary armature with various field excitations. A number of machines were tested at frequencies of 60 and 180 cycles per second and all L/R values were of the same order of magnitude. Figure 6 shows the 60 cycles per second inductance of the machine armature in question for various field excitations.

It is rather difficult to predict the exact effects the variations in armature current have on the armature inductance, the difficulty being caused by there being more than one saturable inductance in the armature circuit. Not only is there inductance in the armature windings, but there also is inductance in the interpole which is connected in the armature circuit. Increasing the direct current tends to produce a relocation of flux paths and changes

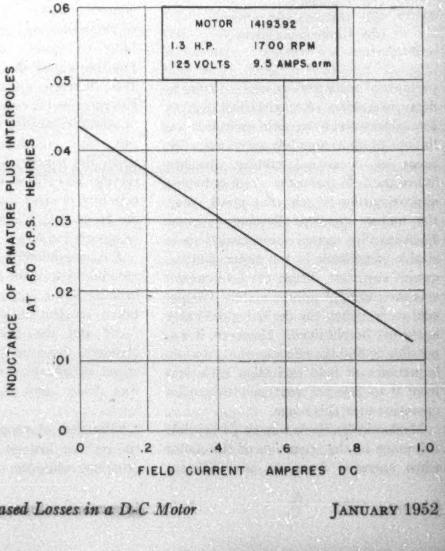


Figure 6. Armature circuit inductance versus shunt field excitation

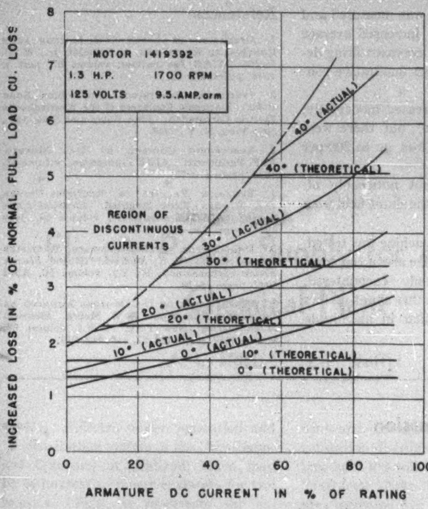


Figure 7. Increase of copper loss due to ripple current

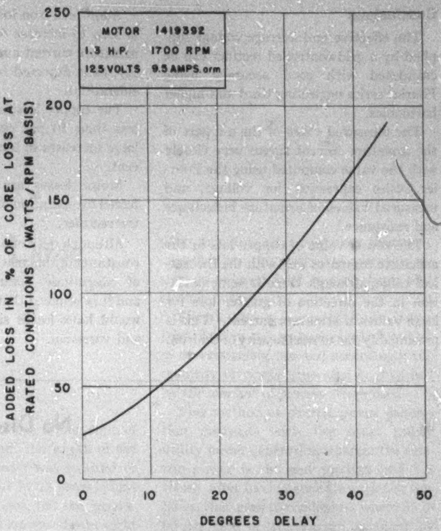


Figure 9. Increase in iron loss versus delay angle for fixed d-c armature current of 75 per cent rated value

in leakage paths. In these calculations it was assumed that the inductances remain constant during change in armature current. Actually, the a-c ripple component increased somewhat as the d-c armature current increased showing that the inductance of the armature circuit decreased with increasing armature current.

As previously stated, the theoretical increase in copper loss is calculated by multiplying the square of the effective ripple component of current by the armature resistance. Assuming that the circuit parameters remain constant during variations in armature current, the theoretical increase in copper loss results in a constant loss for any specified angle of delay. The parameters did not remain constant, however, and the ripple current component increased as the armature current increased. Because of this, copper loss increased as the armature current increased. Figure 7 shows the results obtained from actual tests as well as theoretical values obtained assuming constant ripple current. It can be seen that there is close agreement between the theoretical and actual loss.

Figure 8 shows the increase in iron loss as a function of armature current for various values of rectifier delay. Since the iron loss is affected by the speed of the armature, the data are presented in the form of watts per revolutions per minute in an attempt to eliminate any rotational effects. Although the curves do not present conclusive results, they show a definite trend. The magnitude of the ripple current increases with increased delay as shown in Figure 5. It would, therefore, be expected that the increase in iron loss would also increase with increased delay. Because of the saturation occurring with increased ripple current, the iron loss would increase more rapidly than the ripple current. Figure 9 shows the increase in iron loss per revolutions per minute as a function of rectifier delay with a constant d-c armature current of 75 per cent rated value.

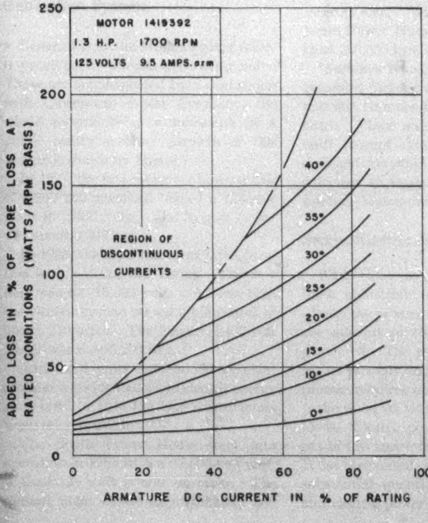


Figure 8. Increase of iron loss due to ripple current, for various delay angles

Conclusions

The effective and average voltage supplied by a grid-controlled rectifier can be calculated with good accuracy using Fourier series neglecting third and higher harmonics.

The measured value of the a-c part of the armature current agrees very closely with the value computed using the Fourier series expression for voltage, and measured values of armature inductance and resistance.

The actual value of copper loss in the armature compares well with the theoretical value, although there is some deviation in the direction of greater loss for large values of armature current. This is presumably due to nonlinearity of the iron.

Additional iron loss was measured and is seen to increase for increased average armature current and increased firing delay, as is expected from qualitative considerations.

The copper loss increased moderately, less than 10 per cent; but there were large increases in iron loss, up to 200 per cent.

Motor losses are not noticeably affected by operation of the shunt field from the rectifier.

Although only one machine was tested, constants of this machine are of the order of magnitude commonly encountered, and it is believed that other shunt motors would have losses similar in magnitude and variation.

References

1. OPERATION OF LARGE D-C MOTORS FROM CONTROLLED RECTIFIERS, **A. Schmidt, Jr., W. P. Smith.** *AIEE Transactions*, volume 67, part I, 1948, pages 679–83.

2. INDUSTRIAL ELECTRONICS REFERENCE BOOK (book), **Electronic Engineers of the Westinghouse Electric Corporation.** John Wiley and Sons, Inc., New York, N. Y., 1948.

3. ELECTRONIC CONTROL OF D-C MOTORS, **K. P. Puchlowski.** *AIEE Transactions*, volume 62, 1943, pages 870–77.

4. BEHAVIOR FACTORS OF RECTIFIER-DRIVEN D-C MOTORS, **Victor Siegfried.** *Electrical Engineering* (*AIEE Transactions*), volume 63, July 1944, pages 530–32.

5. D-C MOTOR OPERATION WITH RECTIFIER POWER SUPPLY, **F. T. DeWolf.** *General Electric Review* (Schenectady, N. Y.), volume 50, April 1947, pages 15–19.

6. PERFORMANCE OF D-C MOTORS RUNNING ON THYRATRON RECTIFIERS, **R. W. Moore.** *Electrical Manufacturing* (New York, N. Y.), volume 37, March, 1947, pages 124–27, 210, 212, 214.

No Discussion

Electric Power Supply for a Large Chemical Plant

A. C. FRIEL
MEMBER AIEE

J. P. SMITH
MEMBER AIEE

THE a-c electric power generation and distribution system of the Dow Chemical Company at Midland, Mich., may be of interest to power engineers for two reasons. First, it represents the experience of a plant wide a-c system that has grown from small beginnings in a period of 23 years and has encountered and solved the problems that go with rapid growth. Second, being a chemical plant, many special problems peculiar to the chemical industry were solved in the process, yielding techniques that may be of interest to others in the chemical and petroleum industry.

Generation System

Generation at the Midland plant today is provided by several generating units. These are supplemented by a connection with Consumers Power Company, the whole scheme being a miniature of a public utility system. Growth of this system is shown by Figure 1.

In 1927, the first major a-c units, a pair of 6,250 kva machines, served a 125-acre plant at 2,400 volts. The longest feeder was about 3,600 feet.

In 1930, the next unit went into operation with a 10,000-kva rating. This new unit was for 13,800 volts. At this time the electric system served a plant area of about 230 acres. The longest 2,400-volt feeder was now 6,200 feet.

Additional units were added as plant requirements for steam and electric energy grew until in 1948 the load was over six times what it was in 1927.

The South Power House went into operation in 1950 with a 37,500-kva unit. A duplicate unit is now on order. The present plant covers approximately 630 acres and through the use of a system of substation distribution, the length of the longest 2,400-volt feeder was reduced to 3,000 feet. Note that in 23 years, plant area increased five times, but the generation kilovolt-amperes increased over eight times.

Two outlying 5,000-kva units complete the present picture of the Dow generating system, these two being located by reason of steam requirements.

The Consumers Power Company supplies power required above our generating capacity which at present is a relatively small portion of the total.

These power sources add up to a total installed capacity for the system of over 150,000 kva. With the next unit at South Power House this will grow to almost 200,000 kva.

Machines Number 6 and 14 are each grounded through a $5^{1}/_{3}$-ohm resistor and machine Number 15 through a 4-ohm resistor. These resistors will keep ground fault current down to 5,000 amperes, sufficient for rapid relay tripping but insufficient to cause excessive burning of machine laminations.

Steam-Electric Balance

With Dow, as presumably it is with most industrials who generate electric power, the process steam demand governs the amount of electric power which is generated. The steam is generated at a higher pressure and temperature than the process requires and is then run through turbines on its way to the process steam loads. Electric power is thereby stripped out of this steam by the turbogenerators. It has been found that where a demand for substantial quantities of process steam and electric power exist within the same plant or within relatively close proximity of each other, this method of operation results in reduced total cost of steam plus electric power.

Electric power generated in an industrial plant in this manner is a by-product of process steam use and accordingly the quantity of power generated is governed by the amount of process steam used.

This method of electric power generation contrasts with the usual public utility power generation wherein the electric power is the end product and the steam, after having gone through the turbines, then goes to condensers where 55 to 60 per cent of the energy supplied by the fuel is lost.

Consideration of these facts makes it apparent that it is desirable under the conditions outlined to maintain as close a balance as possible between steam demands and electric power demands in order to keep requirements for the more expensive purchased power as low as possible.

Figure 1 also shows the growth of electrical load against steam load. The electrical load has developed faster than the steam load and as a result the steam generation temperatures and pressures have been increased and the distribution pressures to process have been reduced to the lowest possible in order to get more electric power out of the steam required by our processes. In spite of considerable effort to increase process steam demand versus power use, purchased power is still required, which is often referred to as condensing power at Dow. Recent studies indicate that Dow could probably make its own condensing power for nearly the same cost as purchased power, but a chemical manufacturer finds it more profitable to invest capital in new process equipment rather than in power-generating equipment on that basis.

Paper 52-80, recommended by the AIEE Industrial Power Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21-25, 1952. Manuscript submitted October 22, 1951; made available for printing December 18, 1951.

A. C. Friel is with the Dow Chemical Company, Midland, Mich., and J. P. Smith is with the General Electric Company, Detroit, Mich.

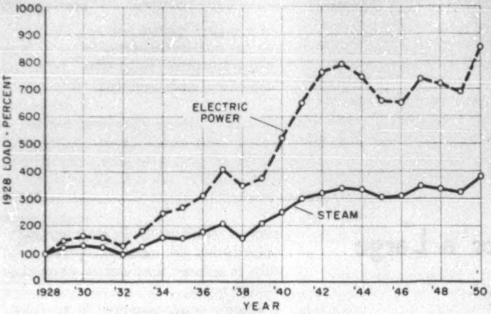


Figure 1. Rate of system load growth

Figure 3. South Power House circuit breaker assembly

With generation of the entire power requirement, Dow would incur the expense and operation of frequency regulating equipment, which is now left to the public utility. Further, it might be considered that purchased power is more firm than the Midland plant generated power since it is backed up by a larger system and therefore, presumably a more stable source. However, the reliability and economy of purchased versus home-generated condensing power is a large subject and one which will not be settled here.

Getting Power to the Load

VOLTAGE SELECTION

The power system at Dow has gone through several stages in its evolution. Originally, 2,400 volts was selected for generation and distribution. Generators were small in size, loads were near by, the power company was able to furnish 2,400-volt power, and this voltage was well suited to motor-generator sets and rotary converter loads.

Later, other units were located farther from the load center. The turbines installed were 9,375 kva, quite large for 2,400 volts. It was at this time that 13,800 volts was selected for all future expansion. This voltage has proven satisfactory with continued plant growth. It today permits a distribution pattern of 7,000- to 10,000-kva loads served by single 3-conductor 500,000 circular mil cables.

SYNCHRONIZING BUS SYSTEM

There is an important factor which is well known but often not considered; short circuit capacity. At one location there are four turbogenerators, but only one machine is connected to a bus section. Three of these generators are 6,250-kva units. With one machine on a bus section, and assuming a 10 per cent reactance generator, the machine would supply 62,500 kva to a bus or feeder fault. Adding the contribution from connected loads and the contribution from ties with current-limiting reactors, 100,000-kva short circuit capacity is very quickly reached.

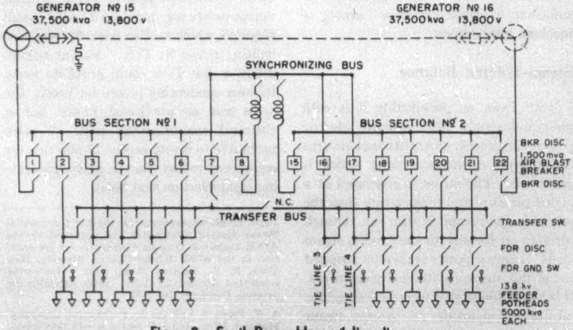


Figure 2. South Power House 1-line diagram

When it is remembered that the largest standard circuit breakers have an interrupting capacity of only 150,000 kva at 2,400 volts, it is obvious that two generators of this size cannot be connected to a common bus, since the short circuit capacity would reach nearly 200,000 kva.

A synchronizing bus system has been adopted. One generator serves each of the four bus sections. A reactor connects each bus section to a main synchronizing bus, which affords means for power transfer between bus sections in the power plant.

Each bus section serves feeders whose normal load will total approximately the normal generated kilovolt-amperes of that bus section. In this way, power transfer through reactors and the synchronizing bus will be kept to a minimum, so that reactor losses are small, and there is very little voltage difference between the various busses.

The synchronizing bus together with changes to be made soon will make it possible to shut down any machine, which may be necessary due to variations in steam demand without interrupting service to any load. For example, if one machine were shut down, power supply to the load on that bus section would not be interrupted, since the three other machines would immediately supply power through their respective bus section reactors to the synchronizing bus, and then through a reactor to the bus section in trouble, or power could be imported from other sources by tie lines. Power never flows through more than two reactors between any generator and any load with this system.

Dow's other powerhouses also use synchronizing bus systems. In one case it has been possible to connect 30,000 kva of generation to one section, and 35,000

Figure 4 (left). South Power House transfer bus

Figure 6 (right). East Side substation number 3

kva of utility power to another section since the 13.8-kv circuit breakers have an interrupting capacity of 600 megavolt-amperes. (Circuit breakers on those sections have recently been rebuilt to provide 960-megavolt-ampere capacity, since new tie lines have raised the short circuit contribution from the synchronizing bus.)

South Power House, with 37,500-kva generators at 13.8 kv, has been provided with 1,500-megavolt-ampere switchgear which will be discussed later.

Experience gained from short circuit studies of this system leads to the conclusion that with a synchronizing bus system, the short circuit capacity of a generator switchgear bus section should be at least 16 times the total name plate kilovolt-amperes of generation connected to that bus (assuming a generator subtransient reactance of approximately 10 per cent).

The synchronizing bus system at the powerhouses also helps to solve a tie-line problem which has been troublesome.

Tie Lines Between Power Stations

Between a pair of Dow powerhouses which shall be called X and Y, a 4,687-kva line ties bus section Number 10 of powerhouse X to section Number 1 of powerhouse Y. A second tie of 8,000-kva capacity connects powerhouse X synchronizing bus with powerhouse Y section Number 1.

It has never been possible to operate these tie lines at rated capacity. At a time when power must be transferred from X to Y, there will be flow through Number 10 section reactor at X. This causes a voltage difference between section Number 10 and the synchronizing bus, the synchronizing bus being at the lower voltage. Thus despite the difference in tie line ratings, they actually carry about the same power, so that actual tie capacity is only 74 per cent of the rated capacity.

The system is being arranged so that both tie lines will terminate at the synchronizing bus at both powerhouses.

There is another pair of ties each of 16,000-kva capacity, terminating at section busses. With present reactances and under certain conditions 70 per cent of the kilovolt-amperes transferred over these ties is carried by line A; 30 per cent by line B. Thus instead of 32,000-kva tie capacity, we have only an effective 23,000 kva. Here also, the system is being arranged with tie lines terminating at station synchronizing busses, which will make the entire 32,000-kva capacity available.

With the system rearranged in this manner, the synchronizing busses and the tie lines will afford a solid backbone for the system. Each section bus will form a practically self-sufficient system with generation equal to load. Upon loss of generation, or under excess load conditions, the power deficiency of any section bus will be made up by power flowing from the synchronizing bus through the section reactor. With this system, there will be only two reactors between any generator and any load in the entire Dow system.

Power House Distribution Busses

At each of the powerhouses, radial feeders carry power directly from the generator bus to the load area. As far as possible, these feeders are so arranged that loads nearest a given powerhouse are

Figure 5. Reactors and bus duct

Figure 7. NT switch house

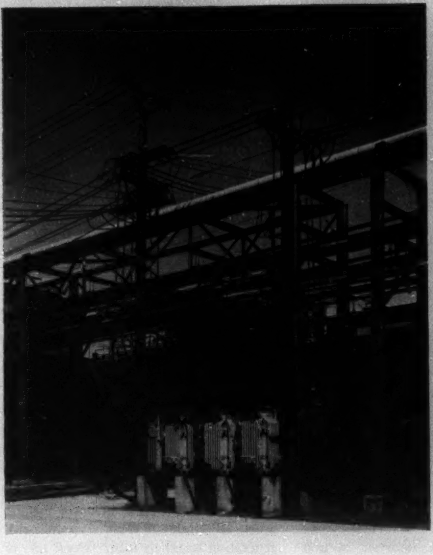


Figure 8. Typical transformer bank

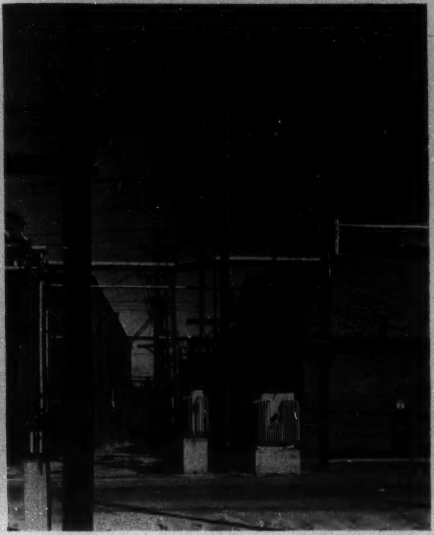


Figure 9. Three-phase transformer

served from that station, and the feeder kilovolt-amperes on a given bus are kept to the normal generated kilovolt-amperes connected to that bus.

In one instance, these two requirements were in conflict. A 5,000-kva substation located near power house *A* would logically be served from that powerhouse, but generation for this load was located at powerhouse *B*. It proved better economics to install a radial feeder the longer distance from powerhouse *B* to this substation, since connecting this load to powerhouse *A* would impose a normal 5,000-kva load on the 32,000-kva tie lines, thus reducing their ability to transfer power between stations during emergencies or machine outages. The switching equipment at the older powerhouses is of open type construction. In order to gain the advantages of indoor protected construction, metal-clad switchgear of 1,500-megavolt-ampere short circuit capacity is used at South Power House, which is the newest one.

Figure 2 shows the manner in which the machines will be connected to bus sections and also how the bus sections are connected through reactors to the synchronizing bus at South Power House. It will be noted that there is a transfer bus which, by proper manipulation of disconnect switches and circuit breakers, can be used to transfer any feeder from its bus section to the transfer bus without dropping load on the feeder. In this way, a feeder circuit breaker can be removed from service for maintenance when required. The transfer bus can also be used to reduce load on a bus section when abnormal conditions require. This arrangement is similar to that used at any of the Dow powerhouses.

Figure 3 shows the installation of 1,500-megavolt-ampere circuit breaker cubicles at South Power House. The addition of new feeders is being accomplished by extending this assembly on the far end. These are station type stationary air blast circuit breakers of 1,500-megavolt-ampere capacity in segregated phase metal-clad construction.

Figure 4 shows the rear of the circuit breaker cubicles, with operating mechanism for circuit breaker disconnects, and on the opposite side of the aisle the transfer bus, with its disconnecting switch mechanisms.

Figure 5 shows the floor below the circuit breaker. The bus duct connecting the circuit breakers with the transfer bus are on the ceiling of this room. The synchronizing bus reactors and the synchronizing bus are also on this floor. Outgoing cables enter potheads in the cubicles at the far side of the room.

It is interesting to note that each feeder circuit breaker will supply two circuits extending to different plant areas. The normal load on each feeder cable is 5,000 kva, so this will permit each circuit breaker to serve a load of approximately 10,000 kva. With a total generation of 37,500 kva per bus section, four feeder circuit breakers are necessary for each bus section. The feeders are so arranged that upon opening of any feeder circuit breaker, the entire 10,000-kva load can be served from other sources as described in the next discussion.

Power Distribution System

Our plant distribution system as it stands today and as it is expected to be for years to come is radial. The plant electrical load is of two types: power for electrolytic service, (largely chlorine production) and power for plant process drives, lighting, and other purposes. The two types of load are approximately equal in kilovolt-ampere demand.

The electrolytic loads are served at 13,800 volts by radial feeders directly, from power house bus sections. This includes rectifiers, motor-generator set substations, and rotary converter substations.

Plant and process load was originally fed by 2,400-volt circuits from one powerhouse but as loads became heavier and feeders become longer, a system of substation distribution was adopted.

A 500,000 circular mill aerial cable at 13,800 volts carries power from the powerhouse to the center of a particular load area. A 2,400-volt load center is installed at this point, consisting of a 6,000/-7,500-kva 3-phase oil filled outdoor transfer bank connected to an indoor 2,400-volt switchgear assembly. Figure 6 shows substation Number 3.

A comparison of cost in 1950 between an indoor and outdoor substation showed that costs ran $15,000 per circuit position for the indoor substation and $14,500 per circuit position for the outdoor substation. It was decided that in view of the relatively severe winter climate, and the chemicals in the atmosphere the indoor substation with its increased operating convenience and reduced maintenance requirements would definitely justify the increased first cost.

Several of these substations are now in service, the first having been installed in 1942, and several more are now in the process of being installed. Working load per substation will be kept down to 5,000 kva so there will be margin for load growth that might occur faster than we anticipate, and also to meet emergency conditions which may arise.

Low-Voltage Ties

The system as explained so far is a purely radial system. However, since many of the processes served from the 2,400-volt substations are critical and cannot be interrupted, a secondary selective feature is provided by normally open 2,400-volt ties between pairs of substations. By sizing the 13.8-kv cable for 10,000 kva, and by using transformers rated 7,500 kva continuous with fan cooling provision has been made for emergency operation of two substations from a single 13.8-kv feeder and transformer. Thus, in case one substation is lost due to a primary failure, it can be served for a short time at full capacity, or for an extended period at reduced capacity, by closing out the secondary tie.

In setting up these secondary ties one of the limitations of 2,400-volt distribution must be faced. The transformer circuit breakers are rated 2,000 amperes, the largest draw-out type 2,400-volt circuit breaker available. This places a name plate ceiling of 8,000 kva on the input to a substation whereas all other equipment and cable is sized for 10,000 kva under emergency conditions. Were 4,160 volts being used, this limitation would not exist.

Use of 4,160-volt distribution in certain areas has been considered, but a change at this time would lead to much confusion, would require rearrangement of circuits, and might require rewinding of some 2,300-volt machines.

It is interesting to compare the old with the new. Figure 7 shows the type of construction used at an old 2,400-volt switch house, which has main bus sections, synchronizing bus, transfer bus, and reactors. At one time all 2,400-volt feeders originated in this switch house. One feeder was 4,800 feet long and another was 6,100 feet long. As may be suspected, voltage problems arose which led to use of step-type voltage regulators.

Today, with substations serving these circuits, the average 2,400-volt feeder is 950 feet long. Carrying power to the load area at 13.8 kv in oversized cable, keeping load well within the rating of the 6,000-kva substation transformer, and keeping 2,400-volt feeders short all contribute to holding excellent voltage at the load without the use of regulators.

Low-Voltage Distribution

Low-voltage load is at 480 volts, 3-phase and 120/240 volts, single phase. At present most of this load is served by open outdoor transformer banks as shown by Figure 8, consisting of 3 200-kva transformers in closed delta plus a fourth transformer for single-phase load, such as lighting. Single-phase transformers formerly were installed for the classical reason; when a failure occurred on one of the single-phase transformers, it could be cut out of service and 58 per cent of the load could be carried by an open delta bank until the faulted transformer was removed and replaced by a spare.

The Midland plant has recently gone to the use of 3-phase transformers, as shown by Figure 9, and is installing 500-, 750-, and 1,000-kva units for our newest banks. It will be noted that the single-phase transformers are retained for lighting. As can be readily seen, this makes a much cleaner looking and simpler installation. And where installation is simpler, it is pretty safe to say that costs will be lower. It is recognized that in case of failure of a 3-phase transformer, the whole load is down while the transformer is removed and replaced by a spare. However, the reliability of a modern transformer is such that the number of failures does not warrant the higher cost of the single-phase transformer construction.

Open Circuit Insulation

The success of our 13,800-volt distribution has been partially due to care in overinsulation of open circuits. The fly ash and chemicals in the atmosphere have led to use of 34.5-kv circuit breaker bushings and bus insulators at the outdoor substation, and for any other open 13,800-volt construction. Even with this insulation, it is necessary to wash insulators and bushings periodically.

The 2,400-volt lines are almost all of open, overhead construction, as shown by Figure 10. These likewise are overinsulated. Where the atmosphere is relatively free from contamination, 15-kv insulators are used; in contaminated at-

Figure 10. Typical 2,400-volt line

Table I. Cost Comparison for Three Types of a 13,800-Volt Plant Power Transmission, Based on 500,000-Circular-Mil Cable

	Max. Kva for 3/C 500 Mcm	Installed Cost per Foot/ Circuit	Cost per Foot/ Mva
One Circuit			
Overhead Open	16,800	$ 7.00	$0.42
Overhead Aerial Cable	10,000	12.50	1.25
Underground Cable in Duct	9,500	15.00	1.58
Four Circuits			
Overhead Open	16,800	3.50	.21
Overhead Aerial Cable	10,000	10.00	1.00
Underground Cable in Duct	7,400	8.50	1.15

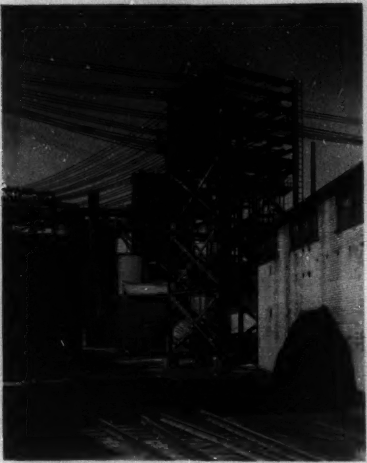


Figure 11 (left). Tower J showing open construction

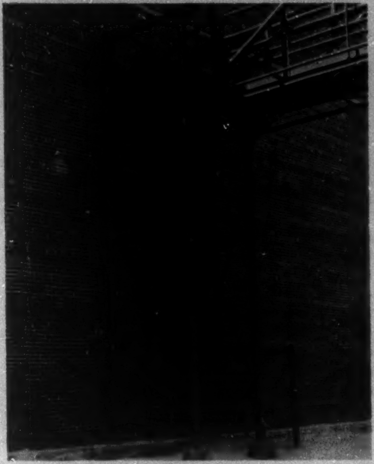


Figure 13 (right). Cable risers at South Power House

mospheres, 22-kv insulators are used.

Figure 11 shows the termination of a group of overhead lines leaving *NT* switchhouse. It may be noted that the lines in tower *J* are considerably over insulated. In the winter of 1942–43 there was a drizzling rain followed by sleet one night, which along with chemical contamination of the atmosphere in this area caused severe flashovers of the insulators on every circuit except one during that night. These lines had always been over insulated as have all other open lines in the plant. However, since that time these particular lines have been doubly over insulated and in addition they are periodically washed hot by means of a fire hose with proper precautions taken against hazard to personnel.

Failures of open 13.8-kv lines has been reduced in recent years by a trick. In the past there were quite a few crossarm and pole fires due to current leakage over the insulators, down the pins to the arms, and across the arms to the poles, thence, presumably to ground. This current would cause heating of the wood and eventually a fire. Bonding together the pins and other hardware on the pole and connecting them to a lag screw driven into the heart of the pole has almost entirely eliminated this problem. The current presumably travels from the bonding circuit down the heart of the pole to ground.

Aerial Cable Versus Open Wire Versus Underground Ducts

The selection of the best line construction for a chemical plant has received much study, and a brief review of the Dow findings may be of interest. In order of expense, open construction is rated lowest in cost, followed by aerial cable, and then by underground duct which is most expensive.

AERIAL INTERLOCKED ARMOR CABLE

The Midland plant originally used only open overhead construction but later turned to aerial cable. The first aerial cable installation was a 15-kv varnished cambric rubber hose jacketed 500,000 circular mil 3-conductor cable with an

Figure 12 (left). Aerial interlocked armor cable

Figure 14 (below). Aerial cable installation

interlocked bronze armor overall. This cable is a generator lead, connecting a 13.8-kv machine which is physically located at one powerhouse to a bus section of another powerhouse, a distance of 2,500 feet. It was installed in 1941, and to date there have been no failures.

It is interesting to note that a lead covered 600-volt multiconductor control cable with performance grade insulation which was installed at the same time as the generator cable and followed the same route failed last spring at several places where cable rings supported it from the messenger wire.

An aerial cable was selected, see Figure 12, since extreme reliability was wanted for this relatively long run. The aerial cable is, of course, not subject to attack by the chemical atmosphere as long as the bronze armor and rubber hose are intact. In addition, aerial cable is less vulnerable to lightning than open overhead lines.

Underground Duct Line

At Midland there is a large amount of underground piping which requires frequent excavation with power shovels. The relatively high water table might present a problem in keeping manholes from flooding. There is a feeling on the part of some of the personnel that the soil in the plant is more or less saturated with chemicals which might have a deleterious effect on the cable. These factors, and the higher cost of underground cable have limited its application to the very few situations where there is no room overhead, where it is desirable to eliminate exposure to lightning, or mechanical damage, or where appearance requirements dictate.

Cost Factors

It is estimated that a 13.8-kv open overhead line of 3 500,000-circular-mil triple-braided-weatherproof wire costs about $7.00 per foot installed on 50-foot poles with spans averaging 100 feet or less. For aerial cable, 3-conductor 500,000-circular-mil cable alone costs about $6.00 to $7.00 per foot, depending on the type of aerial cable used, and installed costs are estimated at $12.00 to $13.00 per foot. These figures are for a pole line carrying one circuit only. Where there is more than one circuit on a pole, the cost per foot, per circuit goes down.

A 500,000-circular-mil aerial cable does not, of course, have the same current carrying capacity as three open triple-braided-weatherproof wires. Therefore the cost per kilovolt-ampere transmitted is still greater with aerial cable.

For an underground circuit of one 500,000-circular-mil 3-conductor cable it is estimated the cost would be about $15.00 per foot installed. A comparison of cost of the three types of construction is shown in Table 1.

Another advantage of aerial cable is saving in space. Figure 11 shows the space required to handle nine 2,400-volt open overhead circuits. Compare this with the space required to get eight circuits away from South Power House, Figure 13. Incidentally, all of the 13.8-kv circuits taking power away from South Power House are 3-conductor, varnished cambric, interlocked bronze armored aerial cable.

A few words on the reliability of aerial cable may be of interest. In the plant there are now approximately 25,000 feet of aerial cable. On this cable there have been three failures which gives a record 0.185 failure per mile per year. There is no similar record for open overhead lines but it can be reported that the number of failures for the open overhead lines per foot per year is considerably in excess of that for aerial cable.

Figure 14 shows the 3-conductor 600,000-circular-mil self-supporting aerial cable in which one of the three aerial cable failures occurred. The contractor had some difficulty getting this cable to hang properly from its messenger since the messenger apparently had a tendency to kink and twist. The installation was difficult because of the somewhat tortuous route the cable had to follow. The route is approximately 800 feet long and has about eight right angle bends as well as a few offsets. Although not noticed at the time of installation, it now appears that this cable had been manhandled somewhat during installation.

Dow has since been advised by two different cable manufacturers that an installation over this route would be considered fairly difficult for any self-supporting type of aerial cable of this size. It is believed that the principal difficulty lies in the tendency of the messenger to twist and turn as the cable comes off the reel. It is passed along to the reader as one experience with larger size self-supporting aerial cable, in the belief that anyone contemplating such an installation should make proper allowances in their planning for it.

Experience With Corrosion of Aluminum Alloys

Purely aside from discussion of the Midland generation and distribution system, the following is a discussion of some experience Dow Chemical Company has had with corrosion of aluminum alloys.

In 1946 it was noticed that the metal rings with which the circuit breaker bushings were held in place at the outdoor substation were severely corroded, some to the extent that a bushing could be expected to pop out of the circuit breaker with resulting fireworks any time one was closed.

An analysis of the metal showed it to be an aluminum alloy having about 4.3 per cent copper. Dow was advised that such corrosion could be expected with this high copper content; and that there were several other alloys with only a trace of copper which would reduce considerably the susceptibility of the alloy to corrosion. The rings in question were replaced with rings of low-copper content and no further trouble has been experienced.

Conclusions

In summary, there are several things that have become apparent from Dow's experience at Midland, and they are passed along to other engineers who may have similar problems in planning their power distribution systems.

1. A purely radial system is simple, easy to operate, and dependable.
2. The total name plate rating of generation connected to a bus section should not exceed 1/16 the short circuit interrupting capacity of that bus.
3. Where several bus sections are found necessary, each section should be tied to the system synchronizing bus by reactors of suitable capacity.
4. A system having generation and loads spread over a considerable area needs a "stiff backbone." Dow is going to build one by means of tie lines between station synchronizing busses.
5. The normal load connected to a bus section should approximately equal the generation and other sources on that section, to avoid carrying power through reactors and tie lines during normal conditions. This keeps tie capacity available for handling emergencies.
6. In laying out a station, provide for enough feeders to take away the power generation connected to that bus.
7. Where the atmosphere bears chemicals and fly ash, build indoor substations, metal enclose, over insulate liberally and use aluminum with low copper content.
8. Geography and availability of equipment are important, but for best results following the preceeding rules for good clean electrical system design is usually more important in planning system changes.

No Discussion

The Writing of Closed-Loop Control System Transfer Functions by Inspection

J. B. FLANNIGAN
NONMEMBER AIEE

H. S. KIRSCHBAUM
ASSOCIATE MEMBER AIEE

IN many closed-loop control systems, an analysis of the system discloses that it is possible for signal energy to traverse different paths and to be modified in a different manner in these various paths. This is in contrast to the very simplest of closed-loop systems, where the signal traverses transfer elements in cascade. The reasons for using multiple-loop systems are many. Improved linearization, noise reduction, and change in impedance level can be achieved by using auxiliary loops. It is also possible to synthesize certain transfer functions by using additional feed-back paths. A simple example of this, is the integral control which can be obtained by using a properly designed series field in a rotating amplifier (Rototrol). It is also possible, by using multiple loops to feed information from various parts of the system to control elements, in order to improve system performance. Also multiple loops might arise where closed-loop control systems are used as components in larger closed-loop systems. For example, in the use of control systems in flight control, the system is part of a larger system closed through the pilot himself.

In multiloop systems, there are two problems to be considered, one being absolute stability and the other being dynamic performance under conditions of absolute stability. In designing such a system before performance can be considered, it is necessary to define the limits under which the system will be stable. In order to do this it is necessary to be able to arrive at the over-all system transfer function, this being numerically equal to the Laplace transform of the system output for a unit impulse input. By examining the over-all transfer function to ascertain whether it has poles in right half plane, the question of stability can be settled. There are two methods of analysis generally used to determine whether the system is stable. These are, the polar plot of the over-all transfer function and the use of the theory of Hurwitz polynomials on the denominator of the over-all transfer function.

For some time it has been the custom of many persons in dealing with closed-loop control systems to write the over-all system transfer function by direct inspection. However, the ability to do this resolves itself into a matter of experience with the networks involved. In some cases, as a final check, it may be necessary to solve the system formally. A method for writing the direct over-all transfer function for 1- and 2-loop systems, with a suggestion for extension to multiple-loop systems, has been published.[1] However, no general method has been presented for multiple-loop systems.

This paper partially fulfills that need by presenting a method of analysis which is applicable to a very large class of systems. It has the advantage of simplicity, speed, and accuracy when compared with the formal solution of a multiloop system.

For a full discussion of the meaning of transfer function, the reader is referred to the references at the end of the paper.

The *n*-Summation-Point System

The methods described in this paper for writing over-all transfer functions are limited to what may be called an *n*-summation-point system with but one forward path between the input and output stations, as shown in Figure 1. A forward path is a path of cascaded unilateral transfer elements connecting the input and output stations in such a way that the direction of information flow is always from input to output. For example, the path $G_1G_2G_3$ in Figure 2 is a forward path, whereas the path $G_1G_2H_{32}$ is not a forward path. The arrows associated with each block indicate the direction of information flow in each element. As shown in Figure 1, the n-summation-point system has n differentials numbered 1 to n. The function of each differential is to subtract from the output of the preceding forward element the sum of all other inputs to the differential. For example, the lth differential in Figure 1 has fed into it the output of $G_{(l-1)}$ and H_{ll}, $H_{(l+1)l}$, $H_{(l+2)l}$, . . . H_{nl}. The output of the differential is the output of $G_{(l-1)}$ less the sum of the outputs of all the H elements connected to it.

It must be noted that there is no connection (through a transfer element) feeding information from, say, the output of $G_{(l-2)}$ to the lth differential. If this were the case there would be more than one forward path from input to output stations, and the rules developed later would not apply.

Any element which is not an element of the forward path will be called a feed-back element. The transfer functions of the forward elements are denoted by G_l and those of the feed-back elements are denoted by H_{ij}.

The Inverse Method

The inverse method consists of obtaining the terms of $\theta_i(s)/\theta_o(s)$ by taking all possible paths from output to input stations, going against the direction of information flow when going through a forward element and with the direction of information flow when going through a feed-back element. To each such path will

---

Paper 52-81, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted October 23, 1951; made available for printing January 7, 1952.

J. B. Flannigan is with Sperry Gyroscope Company, Great Neck, Long Island, and H. S. Kirschbaum is with Ohio State University, Columbus, Ohio.

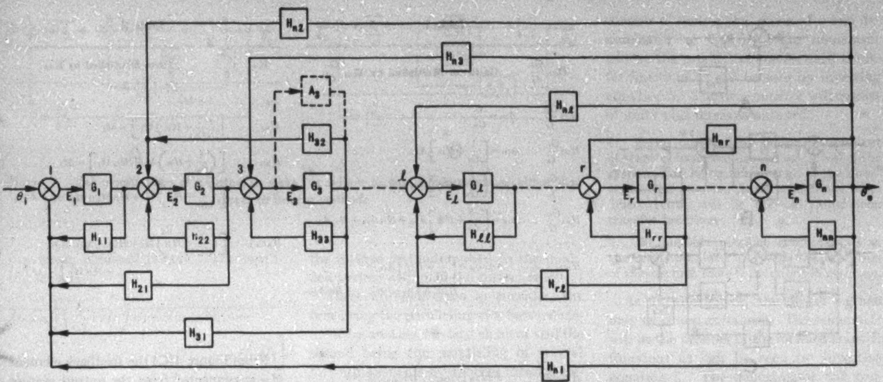


Figure 1. The general n-summation-point closed loop control system

correspond a term of $\theta_i(s)/\theta_o(s)$ and this term will be made up of the product of the feed-back functions divided by the product of the forward functions encountered in the given path. As will be shown later, the over-all inverse transfer function will be the sum of the terms obtained from all the possible allowed paths. For example in Figure 3 there are five possible paths which are allowed. They are as follows, $(G_3G_2G_1)$, H_2, $(H_3G_2H_1)$, $(G_3G_2H_1)$ and $(H_3G_2G_1)$. Hence the inverse transfer function is,

$$\frac{\theta_i(s)}{\theta_o(s)}=\frac{1}{G_1G_2G_3}+H_2+\frac{H_1H_3}{G_2}+\frac{H_1}{G_2G_3}+\frac{H_3}{G_1G_2}$$

from which the direct over-all transfer functions is

$$\frac{\theta_o}{\theta_i}=\frac{G_1G_2G_3}{1+H_2G_1G_2G_3+H_1H_3G_1G_3+H_1G_1+H_3G_3}$$

These are the same results which would be obtained by a formal solution of the system of Figure 3.

Derivation of the Inverse Method

Using the notation of Figure 1, the following set of linear equations can be written, where E_n is the output of the nth differential.

$$\left.\begin{array}{l}
E_1=\theta_i-\sum\limits_{k=1}^{n}G_kH_{k1}E_k\\
E_2=G_1E_1-\sum\limits_{k=2}^{n}G_kH_{k2}E_k\\
\cdots\cdots\cdots\cdots\cdots\\
E_l=G_{l-1}E_{l-1}-\sum\limits_{k=l}^{n}G_kH_{kl}E_k\\
\cdots\cdots\cdots\cdots\cdots\\
\theta_o=G_nE_n
\end{array}\right\}\qquad(1)$$

These are $(n+1)$ linear equations from which the ratio (θ_i/θ_o) can be solved. These equations can be rewritten in the following manner.

$$\left.\begin{array}{l}
(1+G_1H_{11})E_1+G_2H_{21}E_2+\ldots+\\
\qquad\qquad G_nH_{n1}E_n=\theta_i\\
-G_1E_1+(1+G_2H_{22})E_2+G_3H_{32}E_3+\\
\qquad\qquad \ldots+G_nH_{n2}E_n=0\\
0-G_2E_2+(1+G_3H_{33})E_3+\ldots+\\
\qquad\qquad G_nH_{n3}E_n=0\\
\cdots\cdots\cdots\cdots\cdots\cdots\\
0+0+0\ldots+0-G_{n-1}E_{n-1}+\\
\qquad\qquad (1+G_nH_{nn})E_n=0
\end{array}\right\}\qquad(2)$$

which when solved for E_n, yield,

$$E_n=\frac{\left|\begin{array}{ccc}
(1+G_1H_{11}) & G_2H_{22} & \theta_i\\
-G_1 & (1+G_2H_{22})\;\ldots & 0\\
0 & -G_2\quad\ldots & 0\\
\ldots & \ldots\ldots\ldots & 0\\
0 & 0\quad\ldots & 0
\end{array}\right|}{D}\qquad(3)$$

where D is the determinant of the set of co-efficients of equation 2. Hence,

$$E_n=\frac{(G_1G_2G_3\ldots G_{n-1})\theta_i}{D}=\frac{\theta_o}{G_n}\qquad(4)$$

From equation 4 the inverse transfer function is

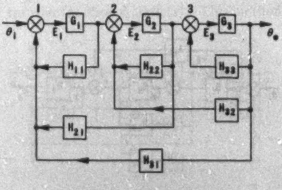


Figure 2. A 3-summation-point system showing all possible feed-back paths

$$\frac{\theta_i}{\theta_o}=\frac{D}{G_1G_2G_3\ldots G_{n-1}G_n}=\frac{D}{G}\qquad(5)$$

where,

$$G=(G_1G_2G_3\ldots G_n)\qquad(6)$$

Before expanding the system determinant, it will be written as equation 7 below.

$$D=\left|\begin{array}{ccccc}
(1+G_1H_{11}) & G_2H_{21} & G_3H_{31} & \ldots & G_nH_{n1}\\
-G_1 & (1+G_2H_{22}) & G_3H_{32} & \ldots & G_nH_{n2}\\
0 & -G_2 & (1+G_3H_{33}) & . & G_nH_{n3}\\
\ldots & \ldots & \ldots & \ldots & \ldots\\
0 & 0 & 0 & \ldots & -G_{n-1}\;(1+G_nH_{nn})
\end{array}\right|\qquad(7)$$

Now it is convenient to expand D upon the nth column obtaining for the inverse transfer function,

$$\frac{\theta_i}{\theta_o}=\frac{G_n}{G}(H_{n1}A_{1n}+H_{n2}A_{2n}+\ldots+H_{nn}A_{nn})+\frac{A_{nn}}{G}\qquad(8)$$

where A_{ln} is the co-factor of the element in row l and column n. The co-factor A_{1n} is obviously equal to G/G_n. Thus the first term of equation 8 will become H_{n1}, which is the term corresponding to the path from output station to input station through the feedback element H_{n1}. The

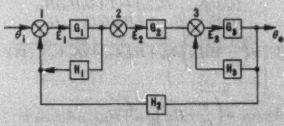


Figure 3. A 3-summation-point system used to illustrate the inverse method

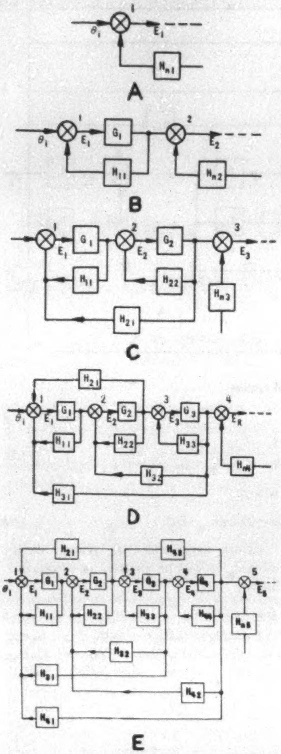


Figure 4. An illustration of the various paths used in the application of the inverse method to an n-summation-point system

remaining co-factors are listed here

$$A_{2n}=(1+G_1H_{11})(G_2G_3\ldots G_{n-1})=\left(\frac{1}{G_1}+H_{11}\right)A_{1n}\qquad(9)$$

$$A_{3n}=(1+G_1H_{11})(1+G_2H_{22})(G_3G_4\ldots G_{n-1})+H_{21}(G_1G_2\ldots G_{n-1})$$
$$=\left(\frac{1}{G_2}+H_{22}\right)A_{2n}+H_{21}A_{1n}\qquad(10)$$

$$A_{4n}=(1+G_1H_{11})(1+G_2H_{22})(1+G_3H_{33})(G_4G_5\ldots G_{n-1})+(1+G_1H_{33})H_{21}(G_1G_2G_4\ldots G_{n-1})+(1+G_1H_{11})H_{32}(G_2G_3\ldots G_{n-1})+H_{31}(G_1G_2\ldots G_{n-1})$$
$$=\left(\frac{1}{G_3}+H_{33}\right)A_{3n}+H_{32}A_{2n}+H_{31}A_{1n}\qquad(11)$$

. . . .

$$A_{(l+1)n}=\left(\frac{1}{G_l}+H_{ll}\right)A_{ln}+H_{l(l-1)}A_{(l-1)n}+\ldots+H_{l1}A_{1n}\qquad(12)$$

. . . .

Table I

$H_{nr}\frac{G_n}{G}$	Co-factor Multiplied by $H_{nr}\frac{G_n}{G}$
$H_{n1}\frac{G_n}{G}$	$A_{1n}=\frac{G}{G_n}$
$H_{n2}\frac{G_n}{G}$	$A_{2n}=\left[\frac{1}{G_1}+H_{11}\right]A_{1n}$
$H_{n3}\frac{G_n}{G}$	$A_{3n}=\left[\frac{1}{G_2}+H_{22}\right]A_{2n}+H_{21}A_{1n}$
$H_{n4}\frac{G_n}{G}$	$A_{4n}=\left[\frac{1}{G_3}+H_{33}\right]A_{3n}+H_{32}A_{2n}+H_{31}A_{1n}$
. . . . . .	. . . .
$H_{n(l+1)}\frac{G_n}{G}$	$A_{(l+1)n}=\left[\frac{1}{G_l}+H_{ll}\right]A_{ln}+H_{l(l-1)}A_{(l-1)n}+\ldots+H_{l1}A_{1n}$
. . . . . .	. . . .

$$A_{nn}=\left(\frac{1}{G_{n-1}}+H_{(n-1)(n-1)}\right)A_{(n-1)n}+H_{(n-1)(n-2)}A_{(n-2)n}+\ldots H_{(n-1)1}A_{1n}\qquad(13)$$

These co-factors, multiplied by (H_{nr} G_n/G) for $r=1, 2, 3\ldots n$ respectively, are the first n terms of equation 8. In addition there is the term A_{nn}/G. These first n terms are tabulated in Table I.

From Table I it would be possible to develop directly the rules of the preceding section. Rather than doing this however, the inverse method will be applied to the system of Figure 1 and a table prepared which is the equivalent of Table I. The term A_{nn}/G will be accounted for separately. The terms of θ_i/θ_o containing H_{nr} as a factor can be obtained systematically as outlined in Figure 4. It will be obvious, without formal proof, that all possible allowed paths will be exhausted.

From Figure 4(A), the feed-back element H_{n1} is connected directly from the output station to differential number 1 and thence to the input station. Application of the rules of the inverse method gives as the first term of θ_i/θ_o the term H_{n1}.

From Figure 4(B), the feed-back element H_{n2} is connected from the output station to differential number 2 and thence through two different paths to the input station. Application of the inverse method to these paths gives the term $H_{n2}(1/G_1+H_{11})$.

Table II

H_{nr}	Term Multiplied by H_{nr}
H_{n1}	$1=M_1$
H_{n2}	$\left[\left(\frac{1}{G_1}+H_{11}\right)M_1\right]=M_2$
H_{n3}	$\left[\left(\frac{1}{G_2}+H_{22}\right)M_2+H_{21}M_1\right]=M_3$
H_{n4}	$\left[\left(\frac{1}{G_3}+H_{33}\right)M_3+H_{32}M_2+H_{31}M_1\right]=M_4$
. . . . . .	. . . .
$H_{n(l+1)}$	$\left[\left(\frac{1}{G_l}+H_{ll}\right)M_l+H_{l(l-1)}M_{(l-1)}+\ldots+H_{l1}M_1\right]=M_{(l+1)}$
. . . . . .	. . . .

From Figure 4(C) the feedback element H_{n3} is connected from the output station to differential number 3 and thence through 5 paths to the input station. Application of the inverse method gives the term

$$H_{n3}\left[\left(\frac{1}{G_2}+H_{22}\right)\left(\frac{1}{G_1}+H_{11}\right)+H_{21}\right]$$

In a similar manner the terms corresponding to Figures 4(D) and 4(E) are respectively,

$$H_{n4}\left\{\left(\frac{1}{G_3}+H_{33}\right)\left[\left(\frac{1}{G_2}+H_{22}\right)\times\left(\frac{1}{G_1}+H_{11}\right)+H_{21}\right]+H_{32}\left(\frac{1}{G_1}+H_{11}\right)+H_{31}\right\}$$

and,

$$H_{n5}\left[\left(\frac{1}{G_4}+H_{44}\right)\left\{\left(\frac{1}{G_3}+H_{33}\right)\times\left[\left(\frac{1}{G_2}+H_{22}\right)\left(\frac{1}{G_1}+H_{11}\right)+H_{21}\right]+H_{32}\left(\frac{1}{G_1}+H_{11}\right)+H_{31}\right\}+H_{43}\left[\left(\frac{1}{G_2}+H_{22}\right)\left(\frac{1}{G_1}+H_{11}\right)+H_{21}\right]+H_{42}\left(\frac{1}{G_1}+H_{11}\right)+H_{41}\right]$$

If the factors multiplying H_{n1}, H_{n2}, H_{n3}, . . . et cetera are denoted respectively by M_1, M_2, M_3, . . . et cetera, then it is obvious that the factor multiplying $H_{n(l+1)}$ namely $M_{(l+1)}$ can be obtained from the

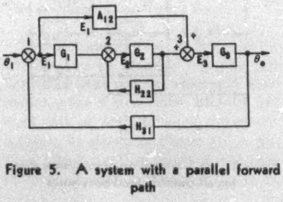


Figure 5. A system with a parallel forward path

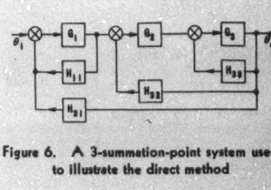


Figure 6. A 3-summation-point system used to illustrate the direct method

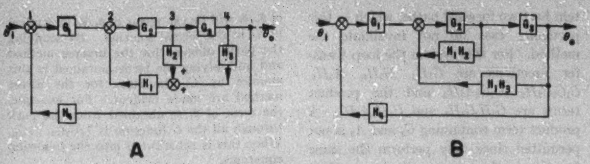


Figure 7. The modification of a system to make it conform to the requirements of either the direct or inverse methods

previously obtained factors. The resulting term is,

$$H_{n(l+1)}\left[\left(\frac{1}{G_l}+H_{ll}\right)M_l+H_{l(l-1)}M_{l-1}+\ldots+H_{l1}M_1\right]$$

All of these terms obtained by the inverse method are tabulated in Table II. A comparison of Tables I and II shows that they are identical insofar as the resulting terms are concerned. These tables show all the terms which are multiplied by H_{rn} for $r=1, 2, 3 \ldots, n$. In addition there is a term obtained from the inverse method by virtue of the path from output station to differential number n through the forward element G_n. Here $1/G_n$ is multiplied by the same factor which multiplies H_{nn} namely M_n. The additional term is thus M_n/G_n. This is the same as the term A_{nn}/G of equation 8. This proves the validity of the inverse method for an n summation point system as shown in Figure 1. Furthermore the solution is in such a form that elimination of any feed-back path or paths will automatically eliminate the same terms in either solution. In the formal solution elimination of H_{rl} corresponds to setting H_{rl} to zero, thus eliminating all terms containing H_{rl}. This corresponds to the result obtained from the inverse method, all other terms remaining intact. The inverse method is valid for this network or any modification of it through the elimination of feed-back paths or the setting of particular G and H functions equal to unity.

Parallel Forward Paths

A type of path which is not present in the network, but which may exist in a practical system, is a forward path in parallel with a portion of the existing forward path of Figure 1. If such is the case the inverse method will not yield the correct result. In some cases the system can be modified in such a way that the modified system is equivalent to the original system and does satisfy the requirement of having only one forward path. Then the inverse method applied to the modified system will yield the correct result.

There are two cases to consider, the first being the paralleling of a forward element by another forward element and the second being the paralleling of several forward elements by another forward element. If in Figure 1 the element G_3 is paralleled by an element A_{33}, the formal solution is still correct if G_3 is replaced everywhere by (G_3+A_{33}), giving rise to such terms as $H_{n4}H_{21}/(G_3+A_{33})$. The inverse method will also work if the elements G_3 and A_{33} are replaced by a single element having a transfer function (G_3+A_{33}). However, if the parallel paths are not reduced in this manner, the inverse method will give two terms $H_{n4}H_{21}/G_3+H_{n4}H_{21}/A_{33}$. Obviously these do not reduce to the correct term. It can be shown that the inverse method, applied to a system with a forward path A_{rl}, will also fail. A particular example will suffice to show this. The formal solution of the system in Figure 5 is,

$$\frac{\theta_i}{\theta_o}=H_{31}+\frac{1+G_2H_{22}}{G_1G_2G_3+G_3A_{12}+G_2G_3H_{22}A_{12}}$$

where A_{12} is a forward element paralleling the elements G_1 and G_2 in cascade. Application of the inverse method to this system would yield a term such as $1/G_3A_{12}$ which does not appear in the formal solution. Hence the inverse method fails in this case.

The Direct Method

It is also possible to derive in a similar manner a method for writing by direct inspection the direct over-all system transfer function $\theta_o(s)/\theta_i(s)$. This method consists of finding the numerator and denominator of $\theta_o/\theta_i(s)$. The numerator will be the product of the forward transfer functions as can be seen by inverting equation 5. The denominator will consist of unity plus terms as outlined:

1. Terms obtained by taking the product of transfer functions around each loop of the system, care being taken to go in the direction of information flow through the loop. These terms will be the individual loop transfer functions.
2. All possible product combinations of the terms obtained in 1 with the exception of those which have G functions in common.

As an alternative, the direct method may be stated as follows. The numerator will be the product of the forward transfer functions as can be seen by inverting equation 5. The denominator will consist of a term for each allowed path as described under the inverse method. Each such term will consist of the product of the H functions encountered with the product of the G functions avoided. For the path through all the G functions in cascade, the term in the denominator is unity.

For example, for the system shown in Figure 6, there are four feed-back loop transfer functions, G_1H_{11}, G_3H_{33}, $G_2G_3H_{32}$ and $G_1G_2G_3H_{31}$. In addition two product terms are allowed, $G_1G_3H_{11}H_{33}$ and $G_1G_3G_2H_{11}H_{32}$. Therefore,

$$\frac{\theta_o}{\theta_i}=\frac{G_1G_2G_3}{1+G_1H_{11}+G_3H_{33}+G_2G_3H_{32}+G_1G_2G_3H_{31}+G_1G_3H_{11}H_{33}+G_1G_2G_3H_{11}H_{32}}$$

or,

$$\frac{\theta_i}{\theta_o}=\frac{1}{G_1G_2G_3}+\frac{H_{11}}{G_2G_3}+\frac{H_{33}}{G_1G_2}+\frac{H_{32}}{G_1}+H_{31}+\frac{H_{11}H_{33}}{G_2}+H_{11}H_{32}$$

which is the result which would be obtained by the inverse method. A proof of the direct method is given in the Appendix.

Further Examples

Throughout the paper examples have been given of the use of both the inverse and direct methods of writing over-a

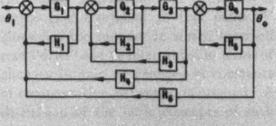


Figure 8. A 4-summation-point system used to illustrate the inverse method

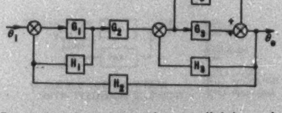


Figure 9. A system with a parallel forward path used to illustrate the direct method

transfer functions by inspection. Several further examples will be dealt with to illustrate particular points of interest. Caution should be observed when applying the inverse method to a system such as shown in Figure 7(A). The path through H_3, H_2, G_2 and G_1 giving rise to a term H_3H_2/G_1G_2 is not allowed since this would violate the rule of traveling in the direction of information flow when traversing a feedback element. If the system is redrawn as in Figure 7(B), it is clear that no such path exists.

A further example will be given to illustrate the inverse method. The system of Figure 8 is characterized by having simple forward and feed-back elements, and hence the inverse method is valid. Hence,

$$\frac{\theta_i}{\theta_o}=H_6+H_5H_4+\frac{H_4}{G_4}+H_5H_3H_1+\frac{H_5H_3}{G_1}+\frac{H_3H_1}{G_4}+\frac{H_3}{G_4G_1}+\frac{H_5H_2H_1}{G_3}+\frac{H_5H_2}{G_3G_1}+\frac{H_5H_1}{G_3G_2}+\frac{H_5}{G_3G_2G_1}+\frac{H_2H_1}{G_4G_3}+\frac{H_2}{G_4G_3G_1}+\frac{H_1}{G_4G_3G_2}+\frac{1}{G_4G_3G_2G_1}$$

To insure that no terms have been omitted it is convenient to analyze the system in the following manner. Any path leading to point 2 will have two possible paths to the input station. There are two paths between points 3 and 2 and hence four paths from point 3 to the input. Now consider point 4. If from point 4, G_3 is entered there are four possible paths to the input. If from point 4, H_2 is entered there are two paths and if H_4 is entered, one path. In other words, there are seven possible paths from point 4 to the input. Since there are two paths from point 5 to point 5 plus an additional path from point 5 to the input through H_6, the total number of paths from output to input are 15. This is in agreement with the number of terms in the above transfer function.

As a final example, the system of Figure 9 will be analyzed by the direct method. It is to be noted that this system has two forward paths, which in this particular case do not invalidate the method. For this system the loop transfer functions are G_1H_1, G_3H_3, A_3H_3, $G_1G_2G_3H_2$, $G_1G_2A_3H_2$ and the product terms are $G_1H_1G_3H_3$ and $G_1H_1A_3H_3$. A product term containing G_3 and A_3 is not permitted since they perform the same function in the system. Such a product term would violate the rule of not having G functions in common in the product terms. The over-all direct transfer function is

$$\frac{\theta_o}{\theta_i}=\frac{G_1G_2(G_3+A_3)}{1+G_1H_1+G_3H_3+A_3H_3+G_1G_2G_3H_2+G_1G_2A_3H_2+G_1G_3H_1H_3+G_1A_3H_1H_3}$$

It is obvious that much simplification would have resulted if the parallel path had been reduced to an equivalent element before applying the direct method.

Conclusion

A method for writing the over-all transfer function of a closed loop control system has been presented. This method is applicable to those systems which have but one forward path between the input and output stations, or to those systems that can be modified so as to have but one forward path. The method is easy to apply and has the advantage over a formal solution in that errors are less likely to occur. In cases where the method will not apply, it is still possible to apply it to parts of the system and thus effect a considerable reduction in the amount of labor involved in the formal solution of the system.

Appendix I. Proof of the Direct Method

The direct method follows directly from the inverse method. The numerator of θ_o/θ_i has already been shown to be the product of all the G_l functions. Since θ_i/θ_o equals D/G, the terms of the denominator of θ_o/θ_i will be the terms of θ_i/θ_o multiplied by G in the formal solution. By studying the paths allowed for the inverse method and multiplying the terms obtained in this manner by G, the rules for the direct method are made evident. For example, the term of θ_i/θ_o obtained from the path through all the G function is $1/G_1G_2\ldots G_n$. When this is substituted into the following equations,

$$\frac{\theta_o}{\theta_i}=\frac{G}{G\left(\dfrac{\theta_i}{\theta_o}\right)} \tag{14}$$

there is obtained in the denominator the required unity of the direct method.

Also, the term of θ_i/θ_o obtained from the path through $G_n \ldots G_{(l+1)}$, H_{ll},

$G_{(l-1)} \ldots G_1$ is $\dfrac{H_{ll}}{G_1G_2\ldots G_{(l-1)}G_{(l+1)}\ldots G_n}$

When this is multiplied by G and placed in the denominator of equation 14, there is obtained a term G_lH_{ll}. Similarly the term

$$\frac{H_{rl}}{G_1G_2\ldots G_{(r-1)}G_{(l+1)}\ldots G_n} \text{ of } \frac{\theta_i}{\theta_o}$$

gives in the denominator of equation 14 the term $G_lG_{(l+1)}\ldots G_{(r-1)}G_rH_{rl}$. This accounts for all the loop transfer functions in the denominator of θ_o/θ_i. The product terms would be obtained from a term of θ_i/θ_o such as

$$\frac{H_{32}H_{rl}}{G_1G_4\ldots G_{(l-1)}G_{(r+1)}\ldots G_n}$$

This would give in the denominator of equation 14 the term $(G_2G_3H_{32})(G_l\ldots G_rH_r)$ which is the product of two loop transfer functions. In this manner the rules of the direct method can be established.

As in the inverse method, the direct method is only applicable to a system of the type shown in Figure 1 or any modification of it based upon setting particular G and H functions equal to unity or particular H functions equal to zero.

References

1. Theory of Servomechanisms (book), H. M. James, N. B. Nichols, R. S. Phillips. McGraw-Hill Book Company, New York, N. Y., 1947.

2. Transients in Linear Systems (book), M. F. Gardner, J. L. Barnes. John Wiley and Sons Inc., New York, N. Y., 1942.

No Discussion

Synthesis of Closed Loop Systems Using Curvilinear Squares to Predict Root Location

D. W. RUSSELL
ASSOCIATE MEMBER AIEE

C. H. WEAVER
ASSOCIATE MEMBER AIEE

THIS paper introduces a procedure for synthesizing servomechanisms whereby the roots of the characteristic equation can be predicted. Given a specified location in the s plane for the roots, the open-loop transfer function parameters and gain can be calculated to give those roots. Since the roots of the characteristic equation are the exponents in the expression for the transient response, it is thus possible to synthesize a system from the standpoint of desired transient response. Inherent in this procedure is the ability to determine whether such a response is physically possible for a particular system. Throughout this paper the term transient response means transient response to a step input.

In the past, since transient studies were not directly applicable to the problem of synthesis, much attention was given to developing other methods. The most widely used method at the present time is the frequency response approach. This method tests for stability by applying the Nyquist Criterion[1,2] as extended to multiple-loop systems by Cauchy's Theorem. Relative stability is associated with the distance the roots of the characteristic equation are located to the left of the j axis in the s plane. The M_p criterion based on the "nearness" of the open-loop transfer function to the $-1+j0$ point, is used as a measure of degree of stability. The great advantage of the frequency response approach, and the reason for its widespread use, is that a fair idea of system response and what must be done to improve system response is secured without having to actually calculate the transient response.

In recent years, to improve on the frequency response approach, other methods for synthesizing servomechanisms have been proposed. In 1950, W. R. Evans[3] presented his "Root Locus Method." This method is concerned with drawing a locus of values of s upon which the roots of the characteristic equation must fall. In 1951 M. R. Aaron[4] presented a paper entitled "Synthesis of Feedback Control Systems by Means of Pole and Zero Location of the Closed Loop Function." The procedure here is to determine the closed-loop function to give a transient response which meets certain specifications and from this, by algebraic manipulations, determine the open-loop transfer function. Both of these methods represent definite advancements in the art of synthesizing closed-loop systems.

The method developed in this paper is based on curvilinear square plots of the open-loop transfer function. The use of curvilinear square plots in synthesis of servos is not original to this paper.[5,6,7] However, the manner in which specified root "drift" can be interpreted in terms of changes in gain or changes in constants of compensating networks, is original. A discussion of the basic concepts of such changes is given in the next paragraph.

If it is desired that all the roots of a system be to the left of a specified line of infinite length in the s plane, parallel to the j axis, this paper presents a procedure for securing such desired root locations. In the frequency response method, the open-loop transfer function is plotted as s takes on all values on the imaginary axis. Applying the Nyquist Criterion, if the $-1+j0$ point is not encircled by this plot, there are no roots to the right of the imaginary axis. If this plot passes through the $-1+j0$ point, there is a root on the imaginary axis. Now, if in place of scanning up the imaginary axis s takes on all values on a line parallel to the imaginary axis but displaced to the left a specified amount, then, if the resulting plot of the transfer function satisfies the Nyquist Criterion for stability, none of the roots of the characteristic equation are to the right of that line. If this line is moved further left until it coincides with a root or roots of the characteristic equation, then the resulting contour will pass through the $-1+j0$ point. As it moves further left, the $-1+j0$ point is encircled. If the line were moved far enough to the left, then all the roots would show up eventually on a contour which passes through the $-1+j0$ point. Curvilinear square plots are a neat way of carrying out this process. Having determined the location of the roots, a method is developed in this paper for shifting the roots a specified amount.

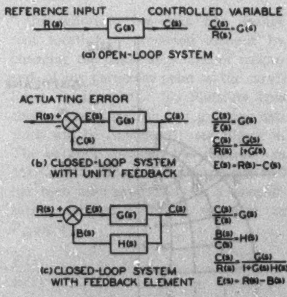


Figure 1. Functional block diagrams

Paper 52-82, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted October 23, 1951; made available for printing December 19, 1951.

D. W. Russell is with ARO, Inc., Tullahoma, Tenn., and C. H. Weaver is with the University of Tennessee, Knoxville, Tenn.

In this paper, only systems which satisfy the following restrictions will be considered. These restrictions are the ones normally applied in servo theory.

1. Only systems that can be described by linear differential equations with constant coefficients will be considered.
2. It will be assumed that there is no energy transfer between blocks.
3. It will be assumed that quiescent initial conditions prevail.
4. The response to external disturbances such as load torque and noise will be neglected.
5. The expression for the transfer function in terms of the Laplace variable s will either be zero or a constant as s approaches infinity.

Background and Discussion

It is now necessary to define some of the terms to be used in this paper. The nomenclature and symbols used are those recommended by the AIEE Subcommittee on Symbols for Feedback Control Systems.[8] Figure 1 illustrates the functional block diagrams used for control systems with the equations relating the variables. Consider the expression relating the output and input for a system with unity feedback

$$\frac{C(s)}{R(s)}=\frac{G(s)}{1+G(s)}$$

$1+G(s)$ is referred to as the characteristic equation of the system. The roots of $1+G(s)$ are the poles of $C(s)/R(s)$ and the zeros of $G(s)$ are the zeros of $C(s)/R(s)$. Roots of $1+G(s)$ are values of s for which $1+G(s)=0$ or $G(s)=-1$. It is this relation that makes studies of the open-loop transfer function useful in analyzing and synthesizing. $G(s)$ is the quotient of two polynomials in s

$$G(s)=\frac{A(s)}{B(s)}=\frac{a_ns^n+a_{n-1}s^{n-1}+\ldots}{b_ms^m+b_{m-1}s^{m-1}+\ldots}$$

where $m>n$

or, in factored form

$$G(s)=\frac{(s+z_1)(s+z_2)(s+z_3)\ldots}{(s+p_1)(s+p_2)(s+p_3)\ldots}$$

where z_1, z_2, and so forth, are zeros of $G(s)$, and p_1, p_2, and so forth, are poles of $G(s)$. The poles and zeros may be complex.

The material in this section will be divided into three major parts. The first part is concerned with the formation of the curvilinear square plots. In the second part, a general expression is derived relating the location of the roots and zeros in the s plane to the coefficients of terms in the expression for the transient response. The third part will take up a commonly used compensating network to show how it can be used to secure predicted improvement in the system response by making predicted changes in the network constants.

CURVILINEAR SQUARE PLOTS

If the loci of s in the s plane form a grid network of squares, then, by conformal transformation, curvilinear squares will be formed in the $G(s)$ plane of constant σ and constant ω contours Figure 2 illustrates this transformation. By Cauchy's theorem,[9] the fact that the curvilinear square *abcd* encloses the point $-1+j0$ means that $1+G(s)$ has zeros or poles or both in the corresponding square *ABCD* in the s plane. By choosing these squares small enough or by using other knowledge about the zeros and poles of $1+G(s)$, it is possible to locate all such zeros and poles.

Alternative to mapping curvilinear squares in the $G(s)$ plane, the inverse, $G^{-1}(s)$, plane can be used; since whenever $G^{-1}(s)=-1$, $1+G(s)/G(s)=1+G^{-1}(s)=0$ and the closed-loop transfer function has a pole. Also, the contours can be transferred to the *Lm*-angle plane though the configuration will not be curvilinear squares.

In many systems, the real roots are located much further to the left in the s plane than the complex roots. For that reason, it may be tedious to find those roots by using curvilinear squares. $Lm\,G(-\sigma)$ versus $\log|\sigma|$ plots afford an easy way of locating those roots. If $G(s)$ is available in factored form, as it usually is in synthesis, the approximate location of each real root can be determined by inspection so that only a few points need be plotted to locate each real root accurately. The real roots occur at values of σ for which $LmG(-\sigma)=0$ decibels. The angle must be -180 degrees or some multiple of 180 degrees.

The actual process of plotting the $G(-\sigma+j\omega)$ contours is very similar to plotting $G(j\omega)$. For example, if $G(s)=K/s(5+s)$, then $G(j\omega)=K/j\omega(5+j\omega)$, and $G(-1+j\omega)=K/(-1+j\omega)\ (4+j\omega)$.

Since, for a constant σ contour, $G(s)$ can be thought of as being a function of ω only, the asymptote methods of making $LmG(j\omega)$ versus $\log\omega$ plots is applicable, though care must be exercised to select proper values of angles and *Lm* at high and low frequencies.

For polar plots, making use of the curvilinear square characteristics of the constant σ and ω contours, it is necessary to calculate only one or two constant σ contours and, remembering that the contours must intersect at right angles, sketch in the constant ω contours and then subdivide each of these squares into smaller curvilinear squares. Sketching between two σ contours can be done quite accurately. The curvilinear square that encloses the $-1+j0$ point can be subdivided further to obtain the value of the pole. Once the plot is sketched in, points of special interest can be calculated to insure accuracy.

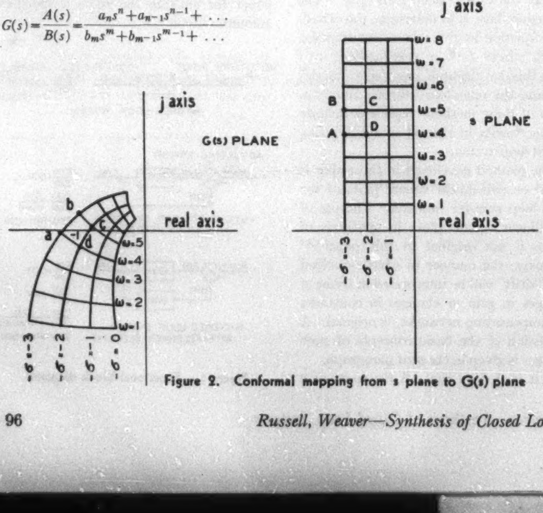


Figure 2. Conformal mapping from s plane to $G(s)$ plane

DERIVATION OF EXPRESSION RELATING LOCATION OF ZEROS AND POLES OF A FUNCTION TO THE COEFFICIENT OF THE TERMS IN THE TRANSIENT RESPONSE

If the location of the zeros and poles of the closed-loop transfer function are plotted in the s plane, it is possible to develop an equation for the general case relating the position of the zeros and poles to the coefficients of the terms in the transient response to a step input.[10] From this equation, some important conclusions can be drawn.

This investigation will be limited to stable systems whose closed-loop transfer functions contain no poles on the j axis and no zeros at the origin. To develop the general equation, it will be assumed that there are only first-order poles. However, the conclusions will be applicable to systems with multiple-order poles. The following relation must also be true, $2n+q>2g+m$ where

m = number of real zeros
g = number of pairs of complex zeros
q = number of real poles
n = number of pairs of complex poles

The closed-loop transfer function can be expressed in operational calculus form as

$$\frac{C(s)}{R(s)}=\frac{G(s)}{1+G(s)}$$
$$=\frac{K\prod_{i=1}^{m}(s+\mu_i)\prod_{i=1}^{g}[(s+\xi_i)^2+\gamma_i^2]}{\prod_{i=1}^{q}(s+\rho_i)\prod_{i=1}^{n}[(s+\alpha_i)^2+\beta_i^2]} \quad (1)$$

Then, if the input $R(s)$ is a step function of magnitude R

$$R(s)=\frac{R}{s} \quad (2)$$

$$\frac{C(s)}{KR}=\frac{\prod_{i=1}^{m}(s+\mu_i)\prod_{i=1}^{g}[(s+\xi_i)^2+\gamma_i^2]}{s\prod_{i=1}^{q}(s+\rho_i)\prod_{i=1}^{n}[(s+\alpha_i)^2+\beta_i^2]} \quad (3)$$

The inverse transformation yields a time response containing a constant term A_o where

$$A_o=\frac{\prod_{i=1}^{m}\mu_i\prod_{i=1}^{g}\gamma_{oi}^2}{\prod_{i=1}^{q}\rho_i\prod_{i=1}^{n}\beta_{oi}^2} \quad (4)$$

Dividing each side of equation 3 by A_o, by inverse transformation, the expression for the normalized time response is

$$\frac{c(t)}{KRA_o}=1+(-1)^{w+1}\sum_{k=1}^{n}M_k\csc\psi_k\epsilon^{-\alpha_k t}\times$$
$$\sin(\beta_k t+\lambda_k+\psi_k)+(-1)^{w}\times$$
$$\sum_{k=1}^{q}(-1)^{k+r_k}N_k\epsilon^{-\rho_k t} \quad (5)$$

where

$$M_k=\frac{\prod_{i=1}^{m}\left|\frac{\mu_{ik}}{\mu_i}\right|\prod_{i=1}^{g}\left[\frac{(\gamma_{ika})(\gamma_{ikb})}{(\gamma_{oi})(\gamma_{oi})}\right]}{\prod_{i=1}^{q}\left|\frac{\rho_{ik}}{\rho_i}\right|\prod_{\substack{i=1\\i\neq k}}^{n}\left[\frac{(\beta_{ika})(\beta_{ikb})}{(\beta_{oi})(\beta_{oi})}\right]} \quad (6)$$

$$N_k=\frac{\prod_{i=1}^{m}\left|\frac{(\mu_i-\rho_k)}{\mu_i}\right|\prod_{i=1}^{g}\frac{(\bar{\rho_k i_\rho})^2}{(\gamma_{oi})^2}}{\prod_{\substack{i=1\\i\neq k}}^{q}\left|\frac{(\rho_i-\rho_k)}{\rho_i}\right|\prod_{i=1}^{n}\frac{(\rho_{ki})^2}{(\beta_{oi})^2}} \quad (7)$$

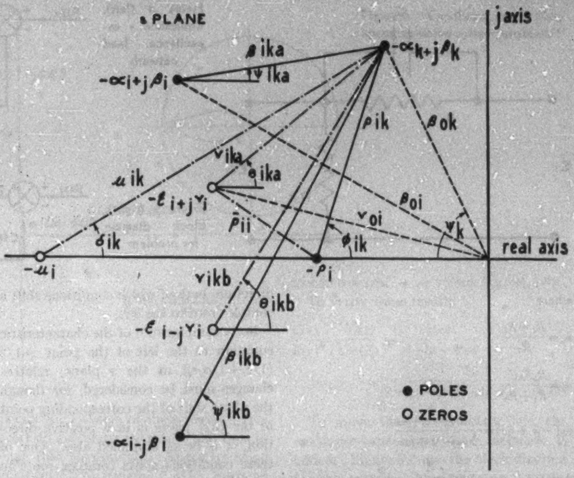


Figure 3. A plot of zeros and poles in s plane

$$\lambda_k=\sum_{i=1}^{g}\theta_{ik}-\sum_{\substack{i=1\\i\neq k}}^{n}\psi_{ik}+\sum_{i=1}^{m}\delta_{ik}-\sum_{i=1}^{q}\phi_{ik} \quad (8)$$

where

w = number of real zeros in right half of s plane
r_k = number of real zeros to right of ρ_k
$(k-1)$ = number of real poles to right of ρ_k

The complete notation is given in Appendix I and is illustrated by Figure 3.

Equations 4, 6, 7, and 8 can be stated as follows:

$$A_o=\frac{\text{product of distances from zeros to origin}}{\text{product of distances from poles to origin}}$$

$$M_k=\frac{\text{product of distances from zeros to } -\alpha_k+j\beta_k}{\text{product of distances from poles to } -\alpha_k+j\beta_k}\times\frac{1/A_o}{(-\alpha_k+j\beta_k)^2}$$

$$N_k=\frac{\text{product of distances from zeros to } \rho_k}{\text{product of distances from poles to } \rho_k}\times\frac{1/A_o}{\rho_k}$$

$$\psi_k=\tan^{-1}\frac{\beta_k}{\alpha_k}$$

λ_k = sum of angles made by lines drawn from zeros to $-\alpha_k+j\beta_k$ minus sum of angles made by lines drawn from poles to $-\alpha_k+j\beta_k$

In calculating M_k, the distance from the conjugate pole should not be included. In calculating ψ_k, the angle made with the conjugate pole should not be included. If there are no zeros, the numerator of A_o, M_k, and N_k is taken as unity.

Examining the statements just made, a number of general conclusions can be made concerning the coefficients of terms in the expression for the transient response.

1. In general, the further the pole is from the origin relative to other poles the smaller its coefficient. This is true because to calculate M_k or N_d it is necessary to divide by the distance from the pole to the origin.
2. A zero near a pole will tend to make the coefficient for that pole relatively small. This condition places a small term in the numerator of M_k or N_k for that pole.
3. If poles are clustered, the coefficients for those poles will tend to be large. This condition places a small term in the denominator of M_k or N_k for those poles.

EFFECT OF COMPENSATION ON ROOT LOCATION

Compensating networks can change root location a predicted amount by changing the phase angle and/or magnitude of any particular point on the curvilinear square plots. To illustrate how this is done, a specific compensating network will be considered; see Figure 4. The network considered is discussed in MacColl's book on servomechanisms and has been used extensively as a compensating network.[11] With reference to Figure 4

$$\frac{E_o(s)}{E_i(s)}=G_1(s)$$
$$=\frac{1/a_1\tau_1+s}{1/\tau_1+s} \quad (9)$$

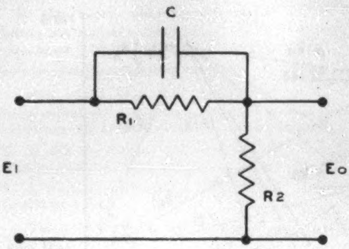
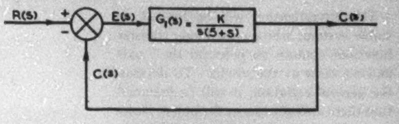
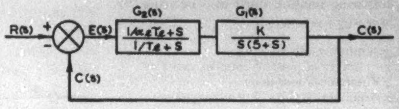


Figure 4 (left). Resistance - capacitance lead network



(a) UNCOMPENSATED SYSTEM



(b) COMPENSATED SYSTEM

Figure 5 (right). Block diagram for problem 1

where

$$\alpha_l = \frac{R_1+R_2}{R_2}$$

$$\tau_i = \frac{R_1 R_2 C}{R_1+R_2}$$

$$G_2(-\sigma+j\omega) = \frac{1/\alpha_l\tau_l - \sigma + j\omega}{1/\tau_l - \sigma + j\omega}$$

$$\text{Ang}\, G_2(-\sigma+j\omega) = \tan^{-1}\frac{\omega}{1/\alpha_l\tau_l-\sigma} - \tan^{-1}\frac{\omega}{1/\tau_l-\sigma} \qquad (10)$$

since $\alpha_l > 1$

$1/\alpha_l < 1$

$1/\alpha_l\tau_l < 1/\tau_l$

$(1/\alpha_l\tau_l - \sigma) < (1/\tau_l - \sigma)$ and

$$\frac{\omega}{1/\alpha_l\tau_l-\sigma} > \frac{\omega}{1/\tau_l-\sigma}$$

therefore

$$\tan^{-1}\frac{\omega}{1/\alpha_l\tau_l-\sigma} > \tan^{-1}\frac{\omega}{1/\tau_l-\sigma}$$

This means that for all values of σ and ω except where $\omega=0$ or ∞. The resistance-capacitance lead network gives a positive phase shift in the $G(s)$ plane. Referring to curvilinear square plots, Figure 2, from the manner in which constant σ contours are traced as the real part of s is made more negative, this positive phase shift will tend to shift complex roots to the left in the s plane.

Now, consider the magnitude of $G_2(-\sigma+j\omega)$.

$$|G_2(-\sigma+j\omega)| = \sqrt{\frac{(1/\alpha_l\tau_l-\sigma)^2+\omega^2}{(1/\tau_l-\sigma)^2+\omega^2}} \qquad (11)$$

for $\sigma \leqq 1/2(1/\tau_l + 1/\alpha_l\tau_l)$

$|G_2(-\sigma+j\omega)| \leqq 1$

for $\sigma \geqq 1/2(1/\tau_l + 1/\alpha_l\tau_l)$

$|G_2(-\sigma+j\omega)| \geqq 1$

For any complex roots of the characteristic equation to the right of the line through the point $-1/2(1/\tau_l+1/\alpha_l\tau_l)$, in the s plane, the root will definitely be moved toward the left, since the corresponding point in the $G(s)$ plane is being attenuated and phase shift is in a positive direction, both of which conditions shift a complex root to the left.

For a complex root of the characteristic equation to the left of the point $-1/2(1/\tau_l+1/\alpha_l\tau_l)$ in the s plane, relative changes must be considered, for though the phase shift of the corresponding point in the $G(s)$ plane is in a positive direction, it is being amplified also. One of these conditions shifts complex roots to the left, and the other shifts them to the right. However, experience has shown that the amplification is quite small and is more than offset by the phase shift. This is true because as σ or ω approach ∞, the magnitude of $G_2(s)$ approaches 1.

Now, consider the effect of the compensating network on the real roots.

$$G_2(-\sigma) = \frac{1/\alpha_l\tau_l - \sigma}{1/\tau_l - \sigma}$$

If $(1/\alpha_l\tau_l - \sigma) > 1$, there is no phase shift.
If $(1/\alpha_l\tau_l - \sigma) < 1$, and
If $(1/\tau_l - \sigma) > 1$, there is a 180-degree phase shift.
If $(1/\tau_l - \sigma) < 1$, there is no phase shift.

Since the real roots must occur at values of σ for which Ang $G_1G_2(-\sigma) = -180$ degrees, and since the "original" real roots occurred at values of σ for which Ang $G_2(-\sigma) = -180$ degrees, the "new" real roots must occur at values of σ for which Ang $G_2(-\sigma) = 0$. The term "new" roots refers to roots of the original system which have been shifted to a new location. Therefore it is necessary to consider only values of σ for which

$$|\sigma| < 1/\alpha_l\tau_l$$

and

$$|\sigma| < 1/\tau_l$$

This means that all the real roots must be between the origin and $-1/\alpha_l\tau_l$ or to the left of $1/\tau_l$. On $LmG(-\sigma)$ versus log $\sigma|$ plots, if the slope of the curve is positive as it crosses the 0-decibel line, then increasing the over-all gain will shift the real root to the right. If the slope of the curve is negative, increasing the over-all gain will shift the real root to the left. In the illustrated problems of the next section, there are several $LmG(-\sigma)$ versus log $|\sigma|$ plots. Referring to these curves, it can be seen that the above statements are true.

Appendix III contains the derivation of an expression to facilitate determination of constants of the resistance-capacitance lead network to give a specified phase shift.

A Procedure for Synthesizing Closed-Loop Servo Systems

The procedure described here is based on moving all the roots of the characteristic equation to the left of a specified line in the s plane. If the magnitude of the real part of each root in the characteristic equation is increased, each term in the expression for the transient response will die out faster resulting in a faster servo.

This procedure involves making curvilinear square plots, determining the roots of the characteristic equation closest to the j axis, and then, after selecting a line in the s plane to the left of which all roots of the system must lie, ascertaining what changes must be made to accomplish this. In most systems, the first roots encountered moving left from the origin in the s plane are a complex pair. For such systems, a new location for the complex pair to improve performance can be specified and, in a straight-forward method, the change in gain or the constants of the compensating network can be determined to move the roots to the specified location.

The procedure can be outlined as follows:

1. Write the open-loop transfer function in factored form.
2. Plot the closure to determine the stability from Nyquist Criterion.
3. Draw the curvilinear square contours in the most convenient plane. As a basis from

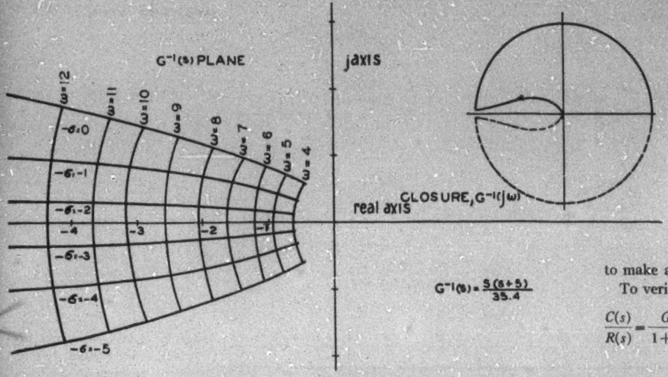


Figure 6. Curvilinear square plot, uncompensated system, problem 1

which to start, the gain may be adjusted for a specified M_p, in the usual manner.

4. From inspection of the open-loop transfer function, determine the approximate location of real roots.

5. Determine the first root encountered moving left from the origin in the s plane. Use curvilinear squares to determine the first complex root, and $LmG(-\sigma)$ versus log $|\sigma|$ plots to determine the first real root.

6. Specify the line in s plane to left of which all roots of characteristic equation must lie.

7. If the first root encountered is a complex root, select a specific location for this root, determine the gain and/or phase shift necessary to move the root to the new location. If the first root encountered is a real root, determine the gain required to move this root left of the line specified in step 6, or near enough to a zero to be neglected.

8. From the phase shift required in step 7, determine the constants of the compensating network.

9. Check physical realizability of the results.

Adding a compensating system to the open-loop transfer function usually increases the number of roots. Frequently, the added root shows up as a real root close to the origin in the left half of the s plane. Such a condition would result in a long tail in the transient response except for the fact that the compensating network also introduces a zero which is in close proximity to this root. This can be seen from equation 9 for the compensating network considered. As is indicated by equation 5, a zero near a root means that the coefficient of that term in the transient response will be small. For a fast response, it is necessary that the coefficient be small.

If the transient response of the system can be approximated by the complex roots nearest the origin, then it is possible to specify the location of these roots to give a certain maximum overshoot and time to reach that maximum overshoot by using equation 17 in Appendix III. Such an approximation is valid for a large number of servos.

Examples Using Procedure Developed in This Paper

To illustrate application of the procedure introduced by this paper two systems are considered—a second-order system and a fourth-order system. The parameters are taken from actual systems. The resistance-capacitance lead network is added to each system.

Problem 1.[12] Compensating a Second-Order System with the Resistance-Capacitance Lead Network

Consider the system of Figure 5. The first step is to plot the closure as illustrated in Figure 6. Then, by frequency response methods, the gain is adjusted for an M_p of 1.3. For this condition, $K=35.4$.

The next step is to plot the curvilinear squares. Arbitrarily, it was decided to make this plot in the inverse plane; in this case, the $G(s)$ plane would have been just as well. The $G^{-1}(j\omega)$ contour ($-\sigma=0$ contour) was plotted very accurately. Then the $G^1(-2+j\omega)$, $G^{-1}(-3+j\omega)$, and $G_1(-5+j\omega)$ contours were plotted. From these constant σ contours, the curvilinear squares were sketched in. The square which encloses the $-1+j0$ point was further subdivided. From Figure 6, the roots were found to be $s=-2.5+j5.37$ and $s=-2.5-j5.37$ since complex roots occur in conjugate pairs. Since this is a second-order system and all the roots have been accounted for, it was not necessary to make a $LmG-(\sigma)$ versus log, $|\sigma|$, plot.

To verify these results

$$\frac{C(s)}{R(s)}=\frac{G(s)}{1+G(s)}=\frac{35.4}{s^2+5s+35.4}$$

$$=\frac{35.4}{(s+2.5+j5.37)(s+2.5-j5.37)}$$

To move these roots to the left, the resistance-capacitance lead network is added. Figure 5 shows the block diagram for this system. Now, suppose it is desired that all the original roots be on or to the left of a line in the complex plane parallel to the imaginary axis and through -4. In order that any gain effects from the resistance-capacitance lead network will not be detrimental and also since it is desirable to have a gain as high as possible with good stability, a specific root can be selected which will require a gain increase. For this problem, it was decided to change the "original" complex roots to $s=-4\pm j8$. From Figure 6, $G_1^{-1}(-4+j8)=2.08\underline{/198.5}$ degrees. If greater accuracy were desired, this specific point could be calculated. Now, $G_1^{-1}(4+j8)G_2^{-1}(-4+j8)$ must equal $1\underline{/180}$ degrees. Therefore, the compensating network must supply a negative phase shift to the inverse function of 18.5 degrees.

Using equation 12 to determine τ_l

$$\alpha_l T^2(1+\tan^2\psi_k)-T\left[\alpha_l+1+\frac{\tan\psi_k(\alpha_l-1)}{\tan\theta}\right]+1=0$$

where

$\tan\psi_k=8/4=2$
$\tan\theta=\tan 18.5 \text{ degrees}=0.335$
$T=\alpha_k\tau_l=4$

since $-\alpha_k=-4$, the real part of the root.

If $\alpha_l=10$, then $\tau_l=0.321$.

Substituting these values for α_l and τ_l into the expression for $G_2^{-1}(s)$,

$$G_2^{-1}(s)=\frac{1/(10)(0.321)+s}{1/(0.321)+s}=\frac{3.12+s}{0.312+s}$$

and

$$G_2^{-1}(-4+j8)=0.904\underline{/-18.5 \text{ degrees}}$$

Figure 7. Curvilinear square plot, compensated system, problem 1

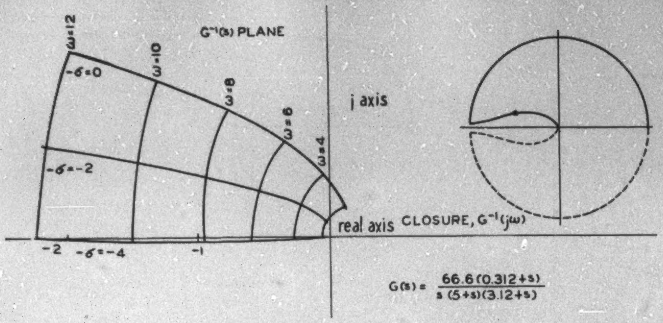


Figure 8. LmG($-\sigma$) versus log $|\sigma|$ plot, compensated system, problem 1

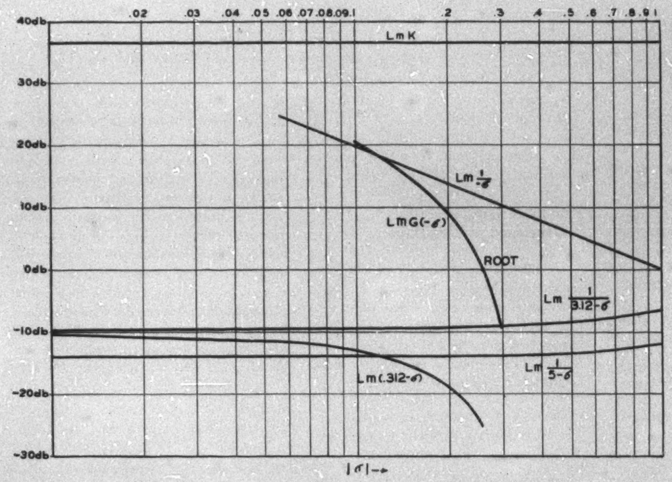
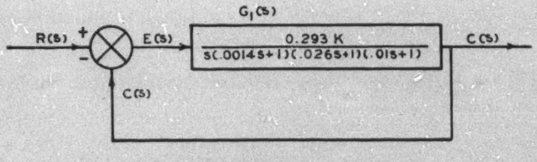
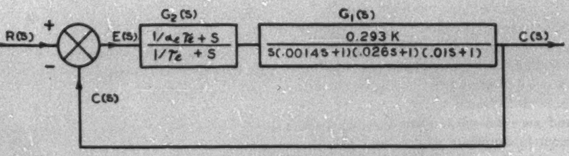


Figure 9. Block diagram for problem 2


G₁(s)
R(s)
+
-
E(s)
0.293 K
s(.0014s+1)(.026s+1)(.01s+1)
C(s)
C(s)

(a) UNCOMPENSATED SYSTEM


R(s)
+
-
E(s)
G₂(s)
1/α₂T₂ + S
1/T₂ + S
G₁(s)
0.293 K
s(.0014s+1)(.026s+1)(.01s+1)
C(s)
C(s)

(b) COMPENSATED SYSTEM

Therefore

$$G_1^{-1}(-4+j8)G_2^{-1}(-4+j8)$$
$$=2.08\underline{/198.5\text{ degrees}}\times$$
$$0.904\,\underline{/18.5\text{ degrees}}=\frac{K/180\text{ degrees}}{35.4}$$

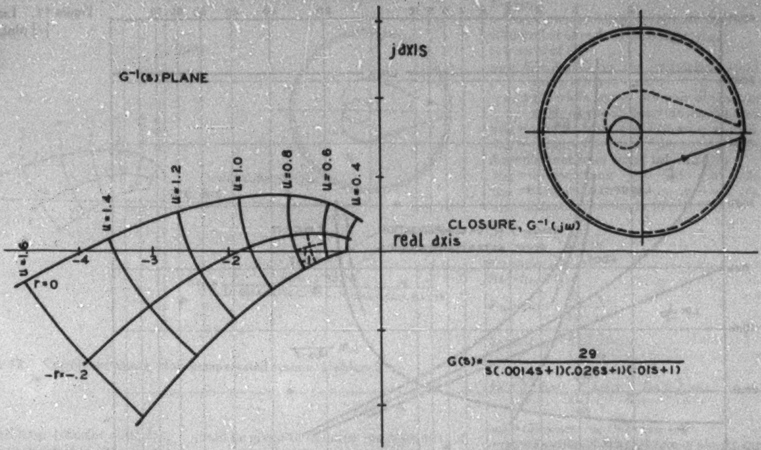
from which $K=66.6$.

The over-all transfer function for the compensated system is

$$G_1(s)G_2(s)=G(s)=\frac{66.6(0.312+s)}{s(5+s)(3.12+s)}$$

This completes the problem, but to show the validity of the process a curvilinear plot was made of the compensated system. This is Figure 7. From this figure, the complex roots are $s=-3.93+$

Figure 10. Curvilinear square plot, uncompensated system, problem 2



$j8.05$ and $s=-3.93-j8.05$, which are within the accuracy expected from a graphical solution.

Adding the compensating network has raised the system to a third-order system. Figure 8 is a $LmG(-\sigma)$ versus $\log|\sigma|$ plot of the compensated system. From this curve, the added root is $s=-0.26$. There is a zero at $s=-0.312$. The proximity of the real root to the zero will cause the coefficient of that term in the transient response to be small. Referring again to Figure 8, notice that the curve crossed the 0-decibel line at a negative slope so that increasing the gain will move this root closer to the zero.

From the manner in which the $G^{-1}(-4+j\omega)$ contour follows the negative real axis, Figure 7, the gain could have been increased considerably without moving the complex roots closer to the j axis. This would have resulted in a much higher K_v.

Verifying the roots

$$\frac{G(s)}{1+G(s)}=\frac{66.6(0.312+s)}{s(5+s)(3.12+s)+66.6}$$

$$=\frac{66.6(0.312+s)}{(s+3.93+j8.05)(s+3.93-j8.05)(s+0.26)}$$

Comparing the system before and after compensation:

Uncompensated System			Compensated System		
K	K$_v$	Roots of Characteristic Equation	K	K$_v$	Roots of Characteristic Equation
3544	7408	$s=-2.5+j5.37$ $s=-2.5-j5.37$	6646	1.33	$s=-3.93+j8.05$ $s=-3.93-j8.05$ *$s=-0.26$

* Near a zero.

PROBLEM 2.[13] COMPENSATING A FOURTH-ORDER SYSTEM WITH RESISTANCE-CAPACITANCE LEAD NETWORK

Consider the system of Figure 9. The first step is to plot the closure as illustrated in Figure 10. Then, by frequency response methods, the gain is adjusted for an M_p of 1.3. For this condition $K=99$. Therefore

$$G_1(s)=\frac{29}{s(0.0014s+1)(0.026s+1)(0.01s+1)}$$

To write this expression in a more usable form, let $s'=r+j\mu=0.0265$ so that $r=0.026\sigma$, $u=0.026\omega$, and dividing by the coefficients on the s terms

$$G_1(s')=\frac{36.5}{s'(18.6+s')(1+s')(2.6+s')}$$

Figure 10 is a curvilinear square plot of $G_1^{-1}(s')$, from which the complex roots are $s'=-0.31\pm j0.76$ or $s=-11.9\pm j29.2$.

Figure 11 is a plot of $LmG_1(-r)$ versus $\log|r|$ from which the real roots are $s'=-2.99$ and $s'=-18.3$ or $s=-115$ and $s=-714$. Notice that these real roots are much further left in the s plane than the complex roots and therefore can be neglected. This is an example of a dominant complex root.

Verifying the values for the roots

$$\frac{G(s)}{1+G(s)}=\frac{29}{s(0.0014s+1)(0.026s+1)(0.01s+1)+29}$$

$$=\frac{79.8\times10^6}{s^4+854s^3+0.1028\times10^6s^2+2.75\times10^6s+79.8\times10^6}$$

$$=\frac{79.8\times10^6}{(s+11.9+j29.2)(s+11.9-j29.2)(s+115)(s+714)}$$

To move these roots to the left, the resistance-capacitance lead network is added. Figure 9 shows the block diagram for this system. Suppose it is desired that all the roots of the original system be on or to the left of a line in the s plane parallel to the j axis and through the point -15.4. To be specific, shift the complex roots so that their value is $s=-15.4\pm j46.2$ or $s'=-0.4\pm j1.2$.

From Figure 10 $G_1^{-1}(-0.4+j1.2)=2.12\underline{/204.3}$ degrees. Since $G_1^{-1}(-0.4+j1.2)G_2^{-1}(-0.4+j1.2)$ must equal $1\underline{/180}$ degrees, the compensating network must supply a negative phase shift to the inverse function of 24.3 degrees. Using equation 12

$$\alpha_l T^2(1+\tan^2\psi_k)-T\left[\alpha_l+\frac{\tan\psi_k(\alpha_{l-1})}{\tan\theta}\right]+1=0$$

where

$$\tan\psi_k=\frac{1.2}{0.4}=3$$

$$\tan\theta=\tan 24.3\text{ degrees}=0.451$$
$$T=r_l\alpha_k=15.4$$

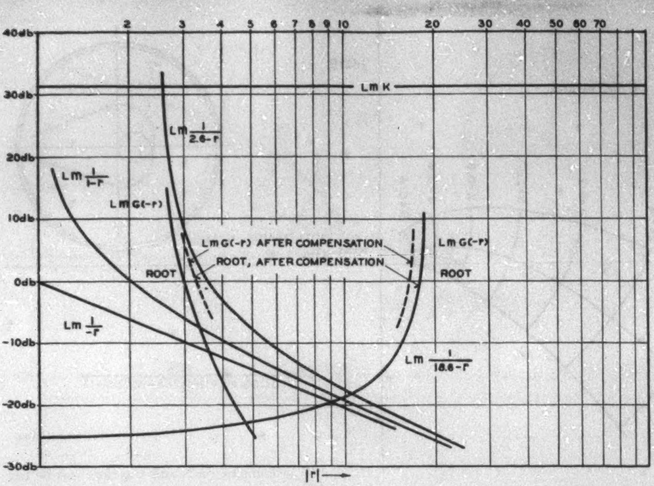


Figure 11. LmG(−r) versus log |r| plot, problem 2

If $\alpha_l = 10$, then $\tau_l = 0.045$.

Substituting these values for α_l and τ_l into the expression for $G_2^{-1}(s)$

$$G_2^{-1}(s) = \frac{22.2+s}{2.22+s} \text{ and } G_2^{-1}(-15.4+j46.2) = 0.97\underline{/-24.3 \text{ degrees}}$$

Therefore

$$G_1^{-1}(-15.4+j46.2)G_2^{-1}(-15.4+j46.2) = \frac{K}{99}\underline{/180 \text{ degrees}}$$

from which $K = 204$.

The over-all transfer function for the compensated system is

$$G_1(s)G_2(s) = G(s) = \frac{59.8(2.22+s)}{s(0.0014S+1)(0.026s+1)(0.01s+1)(22.2+s)}$$

This completes the problem, but, to show the validity of the process, a curvilinear plot was made of the compensated system. This is Figure 12. From this figure, the complex roots are $s' = -0.4 \pm j1.2$ $s = -15.4 \pm j46.2$. Adding $LmG_2(-r)$ versus log $|r|$ plot to $LmG_1(-r)$ versus log $|r|$ plot on Figure 11, and the number of decibels due to the change in gain, the "original" real roots are now $s' = -3.28$ or $s = -126$; and $s' = -18.6$ or $s = -675$.

An additional plot could have been made to determine the value of the root added by the resistance-capacitance lead network. By trial and error, the other real root was found to be $s' = -0.0436$ or $s = -1.68$. Knowing that this root is near the zero, -2.22, it was not an arduous process.

To verify these roots

$$\frac{G(s)}{1+G(s)} = \frac{59.8(2.22s+1)}{s(0.0014s+1)(0.026s+1)(0.01s+1)(22.2+s)+59.8(2.22+s)}$$

$$= \frac{164.4\times10^6(2.22+s)}{s^5+875.5s^4+121{,}810s^3+5.04\times10^6s^2+225\times10^6s+365\times10^6}$$

$$= \frac{164.4\times10^6(2.22+s)}{(s+1.68)(s+126)(s+675)\times(s+15.4+j46.2)(s+15.4-j46.2)}$$

be made to secure a specified change in the location of these roots. It can locate all the roots of the "original system" to the left of a specified line in the s plane. In instances where small real roots are added by compensating elements, it points the way to minimizing the effect of these small real roots much better than frequency response methods. Roots introduced by the resistance-capacitance lead network can be thought of as being in the system originally.

Comparing the system before and after compensation:

Uncompensated System			Compensated System		
K	K_v	Roots of Characteristic Equation	K	K_v	Roots of Characteristic Equation
99	29	$s = -11.9+j29.2$ $s = -11.9-j29.2$ $s = -115$ $s = -714$	204	5.98	$s = -15.4+j46.2$ $s = -15.4-j46.2$ $s = -126$ $s = -675$ *$s = -1.68$

* Near a zero.

Conclusions and Suggestions for Further Study

The procedure for synthesizing servomechanisms developed in this paper is perfectly general, and can be applied to any order system. As was pointed out, it is easier to apply to systems whose first root encountered moving left in the s plane is complex. This method locates all the roots of the characteristic equation and points out exactly what changes must

Consider

$$G_1(s)G_2(s) = G_1(s)\frac{1/\alpha_l\tau_l+s}{1/\tau_l+s}$$

If τ_l is infinite

$$G_1(s)G_2(s) = G_1(s)\frac{s}{s} = G(s)$$

Therefore

$$\frac{G(s)}{1+G(s)} = \frac{G_1(s)s/s}{1+G_1(s)s/s} = \frac{sG_1(s)}{s(1+sG_1(s))}$$

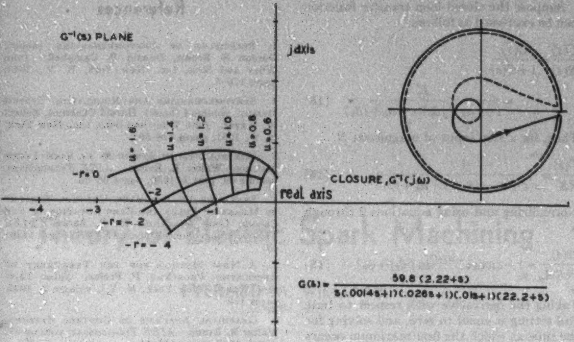


Figure 12. Curvilinear square plot, compensated system, problem 2

so that the closed-loop transfer function has a zero and a pole at the origin. As τ_l becomes finite, this zero and pole at the origin separate and move left.

Possibly the greatest weakness in this method is that nothing has been done about relating the root locations to maximum overshoot and time to reach maximum overshoot except for systems whose response can be approximated by their dominant pair of complex roots. However, as has been stated before, moving roots to the left must result in a system with a faster response. Also, the curvilinear square plots give a direct check with the M_p criterion. In the two examples worked in the preceding section, no attempt was made to shift the roots to result in any specified improvement in system performance. The purpose of the problems was to illustrate a procedure. The value of α_l was selected so that the results could be compared with a problem previously worked by frequency response methods. If a value of 5 has been chosen for α_l, and if K had been increased, a much higher K_v would have resulted. Notice that although increasing gain might cause the complex roots to move toward the right in the s plane, it also causes the small real roots to move closer to the zero, which is desirable.

In the field of feedback amplifiers, J. H. Mulligan, Jr.,[10] has plotted charts relating root location to maximum overshoot and time to reach maximum overshoot. Similar charts especially adapted to servomechanism problems could be plotted for use in conjunction with the procedure introduced in this paper.

In his paper, Aaron derives certain equations relating the real roots to the velocity constant K_v.[4] Possibly attention could be given to defining the constants of the compensating network to move the complex roots a specified amount and then adjusting the gain to give a specified K_v, providing the two specifications are compatible.

To summarize, it is the opinion of the authors that the use of curvilinear squares as applied to this paper can be extended to give even more useful information to the designer. This paper by no means fully exploits the usefulness of this approach. The viewpoint taken in this paper gets to the heart of the problem, the roots of the characteristic equation. It is hoped that it points the way toward further work along these same lines.

Appendix I. List of Symbols

$C(s)$ = controlled variable (output)
$E(s)$ = actuating error
$E_i(s)$ = input voltage
$E_o(s)$ = output voltage
$G(s)$ = over-all open-loop transfer function
$G_1(s)$ = open-loop transfer function before compensation
$G_2(s)$ = compensating block transfer function
$G_1(s)G_2(s) = G(s)$
g = number of pairs of complex zeros in closed-loop transfer function
$H(s)$ = feedback loop transfer function
K = frequency invariant constant
$k-1$ = number of real zeros to right of ρ_k
m = number of real zeros in closed-loop transfer function
n = number of pairs of complex poles in closed-loop transfer function
p_n = poles of $G(s)$
q = number of real poles in closed-loop transfer function
r = nondimensionalized real part of s
r_k = number of real zeros to right of ρ_k
$R(s)$ = reference input
s = Laplace transform variable
$s' = -r + ju$
$T = \tau_l \alpha_k$
u = nondimensionalized imaginary part of s
w = number of real zeros to right of ρ_k
z_n = zeros of $G(s)$
$\alpha_l = (R_1+R_2)/R_2$ in the resistance-capacitance lead network
$-\alpha_k \pm j\beta_k$ = location of kth pair of complex poles of closed-loop transfer function
β_{lka} = distance from $-\alpha_l + j\beta_l$ to $-\alpha_k + j\beta_k$
β_{lkb} = distance from $-\alpha_l - j\beta_l$ to $-\alpha_k + j\beta_k$
β_{ok} = distance from origin to $-\alpha_k \pm j\beta_k$
γ_{lka} = distance from $-\xi_l + j\gamma_l$ to $-\alpha_k + j\beta_k$
γ_{lkb} = distance from $-\xi_l - j\gamma_l$ to $-\alpha_k + j\beta_k$
γ_{ok} = distance from origin to $-\xi_k \pm j\gamma_k$

$$\delta_{lk} = \tan^{-1} \frac{\beta_k}{\mu_l - \alpha_k}$$

θ = angle of $G_2(s)$
$\theta_{lk} = \theta_{lka} + \theta_{lkb}$

$$\theta_{lka} = \tan^{-1} \frac{\beta_k - \gamma_l}{\xi_l - \alpha_k}$$

$$\theta_{lkb} = \tan^{-1} \frac{\beta_k + \gamma_l}{\xi_l - \alpha_k}$$

$$\lambda_k = \sum_{l=1}^{g} \theta_{lk} - \sum_{\substack{l=1 \\ l \neq k}}^{n} \psi_{lk} + \sum_{l=1}^{m} \delta_{lk} - \sum_{l=1}^{q} \phi_{lk}$$

μ_{lk} = distance $-\mu_l$ to $-\alpha_k + j\beta_k$
$-\mu_k$ = location of kth real zero of closed-loop transfer function
$-\xi_k \pm j\gamma_k$ = location of kth pair of complex zeros of closed-loop transfer function
$-\rho_k$ = location of kth real pole of closed-loop transfer function
ρ_{lk} = distance from $-\rho_l$ to $-\alpha_k + j\beta_k$
$\bar{\rho}_{lk}$ = distance from $-\rho_l$ to $-\xi_k + j\gamma_k$
$-\sigma$ = real part of s
$\tau_l = \dfrac{R_1R_2C}{R_1+R_2}$ in the resistance-capacitance lead network

$$\phi_{lk} = \tan^{-1} \frac{\beta_k}{\rho_l - \alpha_k}$$

$\psi_{lk} = \psi_{lka} + \psi_{lkb}$

$$\psi_{lka} = \tan^{-1} \frac{\beta_k - \beta_l}{\alpha_l - \alpha_k}$$

$$\psi_{lkb} = \tan^{-1} \frac{\beta_k + \beta_l}{\alpha_l - \alpha_k}$$

$$\psi_k = \tan^{-1} \frac{\beta_k}{\alpha_k}$$

ω = imaginary part of s

Appendix II. Derivation of Equation Relating Phase Shift and Constants of the Resistance-Capacitance Lead Network

With reference to Figure 4

$$\frac{E_o(s)}{E_i(s)} = G_2(s)$$

$$= \frac{1/\alpha_l\tau_l + s}{1/\tau_l + s}$$

where

$$\tau_l = \frac{R_1R_2C}{R_1+R_2}$$

and

$$\alpha_l = \frac{R_1+R_2}{R_2}$$

from which

$$G_2(-\alpha_k+j\beta_k)=\frac{1/\alpha_l\tau_l-\alpha_k+j\beta_k}{1/\tau_l-\alpha_k+j\beta_k}$$

$$\text{Ang}\, G_2(-\alpha_k+j\beta_k)=\theta$$

$$=\tan^{-1}\frac{\beta_k}{1/\alpha_l\tau_l-\alpha_k}-\tan^{-1}\frac{\beta_k}{1/\tau_l-\alpha_k}$$

which simplifies to the following expression

$$\alpha_l T^2(1+\tan^2\psi_k)-T\left[\alpha_l+1+\frac{\tan\ \psi_k(\alpha_l-1)}{\tan\theta}\right]+1=0 \quad (12)$$

where

$$T=\tau_l\alpha_k$$

$$\psi_k=\tan^{-1}\frac{\beta_k}{\alpha_k}$$

Experience by engineers in industry has shown that for best results

$$5<\alpha_l<20$$

Appendix III. Derivation of the Expression Relating the Maximum Overshoot to the Ratio of the Imaginary to the Real Part of the Dominant Root Pair

Consider the general equation for a second-order system.

Suppose the closed-loop transfer function can be expressed as follows

$$\frac{C(s)}{R(s)}=\frac{G(s)}{1+G(s)}=\frac{K}{(s+\alpha_k-j\beta_k)(s+\alpha_k+j\beta_k)} \quad (13)$$

Then, for a step input of magnitude R

$$\frac{C(s)}{RK}=\frac{1}{s(s+\alpha_k-j\beta_k)(s+\alpha_k+j\beta_k)} \quad (14)$$

Normalizing and using equations 2 through 5

$$\frac{c(t)}{RKA_0}=1-csc\psi_k\epsilon^{-\alpha_k t}\sin(\beta_k t+\psi_k) \quad (15)$$

Taking the derivative with respect to time and setting it equal to zero, and solving for the time at which the first maximum occurs

$$t=\frac{\pi}{\beta_k} \quad (16)$$

Substituting equation 16 into equation 15, and simplifying

$$\left.\frac{c(t)}{RKA_0}\right|_{\max}=1+\epsilon^{-\pi\cot\psi_k} \quad (17)$$

This equation brings out the usefulness of the dominant root approximation. It expresses the maximum overshoot as a function of the single variable ψ_k, where

$$\psi_k=\tan^{-1}\frac{\beta_k}{\alpha_k}$$

References

1. Principles of Servomechanisms (book), **Gordon S. Brown, Donald P. Campbell.** John Wiley and Sons, Inc., New York, N. Y., 1948, pages 170–5.

2. Servomechanisms and Regulating System Design, volume I (book), **Harold Chestnut, Robert W. Mayer.** John Wiley and Sons, Inc., New York, N. Y., 1951, pages 138–56.

3. Control System Synthesis by Root Locus Method, **Walter R. Evans.** *AIEE Transactions*, volume 69, part I, 1950, pages 66–69.

4. Synthesis of Feedback Control Systems by Means of Poles and Zero Location of the Closed Loop Function, **M. R. Aaron.** *AIEE Transactions*, volume 70, part II, 1951, pages 1439–46.

5. A New Method for the Treatment of Regulation Problems, **P. Profos.** *Sulzer Technical Review* (New York, N. Y.), volume 2, 1945, pages 1–5.

6. Graphical Analysis of Control Systems, **Walter R. Evans.** *AIEE Transactions*, volume 67, part I, 1948, pages 547–51.

7. Introduction to Servomechanisms (book), **A. Porter.** John Wiley and Sons, Inc., New York, N. Y., 1950, pages 93–4.

8. *Op. cit.*, **Harold Chestnut, Robert W. Mayer,** pages 195–9.

9. Fundamental Theory of Servomechanisms (book), **Leroy A. MacColl.** D. Van Nostrand Company, Inc., New York, N. Y., 1945, page 23.

10. Effect of Pole and Zero Location on the Transient Response of Linear Dynamic Systems, **J. H. Mulligan,** Jr. *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 37, May 1949, pages 516–29.

11. *Op. cit.*, **Leroy A. MacColl,** page 51.

12. *Op. cit.*, **Gordon S. Brown, Donald P. Campbell,** page 371.

13. *Ibid.*, page 379.

No Discussion

Theory of Electric Spark Machining

EVERARD M. WILLIAMS
MEMBER AIEE

Synopsis: The electric spark machining process is becoming of considerable industrial importance in the tool and die industry. A typical electrical system for this process is described and a theory is advanced for the mechanism of operation. The results of experimental tests are reported.

THE electric spark process for the machining of hard metals is in extensive use abroad and has been introduced recently in the United States.[1] It has considerable industrial interest since it permits, for instance, the rapid cutting of accurate and, if necessary, intricate shapes in such materials as sintered tungsten carbides, *S-816*, vitalium, hardened steels, and hard alloys of titanium, zirconium, vanadium, and so forth which are ordinarily not machinable. This process is now used, for instance, for the finishing of hollow jet turbine blades, turbine blade roots, thread and rifle gauges, nozzles, bodies with re-entrant surfaces, and so forth. This paper describes the basic features of this process and the results of an experimental and theoretical study of the phenomena involved.

In the electric spark machining process, as described by Lazarenko[2-4] and others[5-7] a succession of high current sparks are caused to pass between a work piece and an electrode having the shape of the hole which it is desired to pierce through the work piece. As shown in Figure 1; the sparks are derived from the discharges of a capacitor. Each spark more or less completely discharges the capacitor, following which the spark path is deionized and the capacitor is recharged from the d-c source through the charging resistance. During the process, material is removed from both electrode and work piece and the electrode is advanced manually or automatically[8,9] until the machining operation is complete. Lazarenko has shown that cutting speed is greatly increased if the electrode polarity is negative with respect to the work piece and if the gap between electrode and work piece is immersed in a dielectric, commonly kerosene or transformer oil. Figure 2 shows a complete spark-cutting machine and Figure 3 some samples of tungsten carbide machined by this method.

The physical phenomena associated with the process are as follows. The electrode, connected permanently to the negative lead of the charged capacitor, is advanced through the dielectric toward the work piece which is connected to the positive electrode of the capacitor. When the gap between electrode and work piece is reduced to a sufficiently small value, of the order of 0.003 inch per hundred volts, the dielectric breaks down and a discharge takes place. The range of working voltage in production machines is from about 40 to 400 volts. Oscillographic measurements of typical discharges show that the discharges are oscillatory with a range of peak currents from a few hundred to about 20,000 amperes and frequencies of from 10 to 1,000 kc. Each discharge is accompanied by explosive sounds, spattering of oil, and evolution of gas bubbles, smoke, light, and heat; these effects are particularly pronounced at the higher voltages and currents. Following the discharge, the gap deionizes. It is probable that deionization is assisted by the liberation of hydrogen in the spark region through chemical decomposition of the dielectric fluid, such as occurs in oil circuit breakers. Following deionization, the capacitor recharges and the electrode advances until a voltage is reached at which breakdown again occurs. Typical spark repetition rates range from 50 to 5,000 discharges per second. The detached fragments of the work piece and electrode material are removed by the dielectric flow resulting from the local pressures caused by the explosive discharges. This removal may be augmented by forced circulation of the dielectric fluid. The duty cycle is extremely low due to the low recovery rate of capacitor voltage necessary to avoid arc formation and the time required for removal of eroded material. The holes produced are larger than the electrode by the gap spacing corresponding to the breakdown distance associated with the peak capacitor voltage under any particular operating condition. Variations in the gap due to variations in electrode and work piece surface configurations as machining proceeds result in a rough machined surface. For fine finishing cuts this and other effects are minimized by using very low voltages and capacitances.

Spark machining is slow by comparison with conventional machining methods for soft materials; rates of material removal, for instance, are of the order of 1 gram per minute. The process is commercially feasible for very hard materials, however, as it permits the forming of desirable tool and die shapes unobtainable by alternative means and, where alternative means exist, is often more economical of material and time.

A study has been made of the erosion produced by single sparks in transformer oil between an electrode and a plane tungsten carbide surface under conditions corresponding to fairly rapid and rough machining by the electric spark process. In these tests, peak currents, frequencies of the oscillatory discharge, and gap spacing were varied and the depth and area of the craters produced were studied by means of a microscope. In each case, rough craters were produced characterized by a well defined area of nearly constant depth.

Experimental data, which are the aver-

Paper 52-101, recommended by the AIEE General Industry Applications Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21-25, 1952. Manuscript submitted October 19, 1951; made available for printing December 5, 1951.

EVERARD M. WILLIAMS is with the Carnegie Institute of Technology, Pittsburgh, Pa., and is consultant to Firth Sterling Steel and Carbide Corporation, under whose auspices the work described in this paper was carried out.

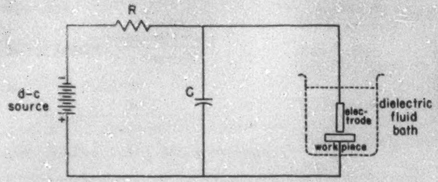


Figure 1. Schematic diagram of electric circuit of spark cutting machine used by Lazarenko

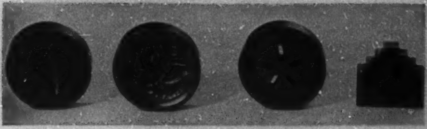


Figure 3. Sintered tungsten carbide pieces formed by spark cutting machine. Left, two engraving dies; second from right, die used in textile manufacture; and, right, step drill tip

age of a large number of observations, are plotted in Figures 4, 5, 6, and 7. Figures 6 and 7 have been plotted with ordinates normalized to show particular relations to be discussed hereafter. It is apparent that crater depth depends on the magnitude of discharge current while crater area is very nearly proportional to discharge duration. Material removed by sparks of varying amplitude and fixed duration is roughly proportional to the three-halves power of peak discharge current. Other measurements show that gap width, corresponding to voltage with which the capacitor is charged at instant discharge takes place, has little effect on crater depth and area, but is significant in that the crater location with respect to electrode position becomes increasingly erratic as the gap width is increased. Operating voltage thus has an important effect on machining accuracy for this reason in addition to its effect on the variations in the critical gap previously discussed.

Electrode material has little effect on the dimensions of the crater produced in the work piece. Erosion of the electrode itself, however, varies considerably between materials and is also a function of the electrode form. Differences in erosion between various electrode materials are very much decreased when thin tubular electrodes are used.

The spark erosion process in which erosion is greatest at the positive electrode is evidently quite different from arc cutting systems in which the greater erosion takes place at the negative electrode. Arc erosion is accompanied by high temperatures and melting, whereas in the spark erosion process microscopic examination of the work piece surface gives evidence of fracture by mechanical forces. Analyses of residues from spark machining of sintered tungsten carbides, for instance, give no evidence of physical or chemical changes that might be associated with high temperatures or melting of the tungsten carbide. Machined surfaces exhibit no alteration in temperature related properties such as surface hardness. Crystallographic studies of machined surfaces of material with transformation temperatures slightly above ambient have shown that machined surface temperature rises are insufficient to produce such crystal transformations. Therefore, it may be concluded definitely that material is removed without melting.

Several theories have been advanced to explain spark erosion. Lazarenko has suggested electrolytic action and "elec-

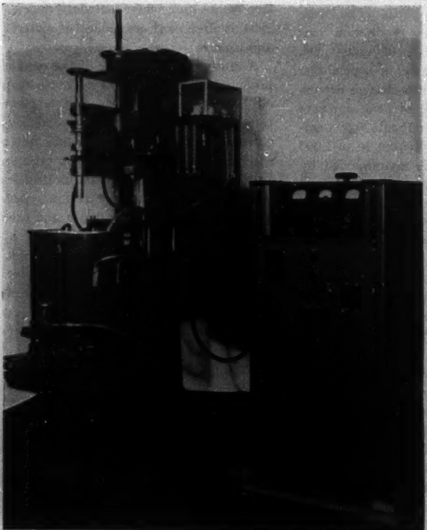
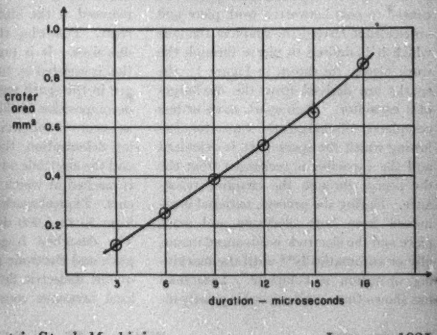


Figure 2 (left). Spark cutting machine of the form now in use in the United States, left, and control cabinet, right. Machine is knee and column type with lateral and longitudinal slides. Fluid tank rests on a 360 degree rotary base. The column carries the work head and the electrode driving motor is at top rear. The control cabinet carries rectifiers and the servomechanism for automatic electrode drive. Once set up a particular machining operation is completely automatic and requires no attention

Figure 4 (below). Area of crater in sintered tungsten carbide resulting from a single capacitor discharge between brass electrode and plane surface of tungsten carbide work piece as a function of half-cycle duration of a 5,600-ampere (peak) oscillatory discharge in microseconds. The negative and positive terminals of the charged capacitor are connected to the electrode and work piece respectively. The oscillatory discharge is sufficiently damped so that only initial half-cycle has appreciable effect on crater



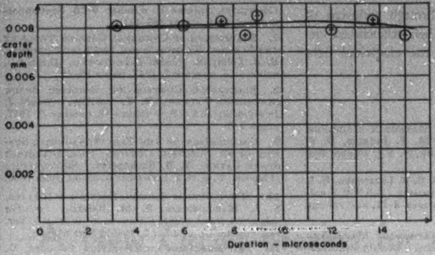


Figure 5. Depth of crater as a function of spark duration under same conditions as those of Figure 4

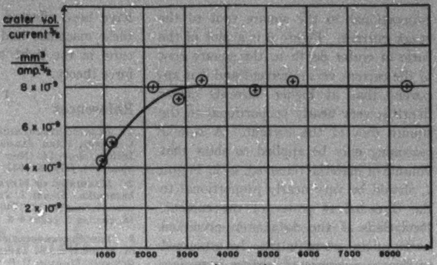


Figure 7. Effect of peak current on volume of crater under conditions of Figure 6 expressed as a function of ratio of crater volume to square root of current cubed, showing dependence of crater volume on square root of current cubed at higher peak currents

trodynamic" forces of an undisclosed nature. Mandelshtam and Raiskii[10] have propounded a theory in which erosion is due to the explosive force of the discharge augmented by a high temperature metal-vapor flare propagated from the cathode (electrode) toward the anode (work piece). Rudorff[11] has suggested that oscillating mechanical forces, similar to those involved in cavitation, are involved. Evidence in support of these theories however, is qualitative in nature and inconclusive.

A very satisfactory quantitative correlation with experimental results has been obtained, however, for a theory which accounts for a major component of the erosion by electrical field forces. These forces arise because of the extremely high current density at, about, and beneath a point on the surface of the work material and a correspondingly high electric field gradient and force on the positive ions of the crystal lattice. The magnitude of the forces involved may be estimated from a knowledge of the crater depth and area produced by a known discharge current. In a typical instance, a current of 5,600 amperes produced in sintered tungsten carbide a crater 0.0085 millimeter deep and 0.078 square millimeter in area. The peak surface current density J_s may be estimated from the peak current and crater area as, at least, $J_s = 5{,}600/0.078 = 71{,}700$ amperes per square millimeter or about 7,170,000 amperes per square centimeter. The resistivity of the tungsten carbide sample is about 28×10^{-6} ohm-centimeter so that the normal component of electric field gradient E_s just beneath the surface is $E_s = 201$ volts per centimeter. The number of free electrons (which is equal to the number of positive ionic charges) in a relatively good conductor such as tungsten carbide, is of the order of magnitude of the number of metal atoms.[12] In tungsten carbide, this criterion indicates a positive charge density of 5.8×10^3 coulombs per cubic centimeter, or in the fragment corresponding to the crater in this case a charge of 3.9×10^{-3} coulomb. The force on the fragment is thus 0.784×10^7 dynes or 17.6 pounds. This is equivalent to a stress at the bottom of the fragment, neglecting shear stresses on its sides, of 145,000 pounds per square inch, which is sufficiently in excess of the breaking strength of the material to account for fracture. According to this theory a larger fragment than that observed is not detached because the normal component of current density decreases below the surface due to spreading of the current, which is undoubtedly augmented by skin effect. Erosion of the electrode material is due to:

1. The oscillatory nature of the spark discharge in which the electrode becomes positive in the second half cycle and
2. Erosion of the soft electrode material by the stream of hard workpiece particles removed from the workpiece.

This electrical field force hypothesis as to the nature of the erosion force provides explanations for the curves of Figures 4 to 7 as follows:

Figures 4 and 5. Crater depth is independent of spark duration while crater area is directly proportional to spark duration because in the course of a single discharge several fragments are detached in sequence from adjacent areas with the entire current of the discharge terminating on each area in turn. The actual current density in the discharge at any time is, therefore, considerably greater than that estimated from the total crater diameter. Since the depth of the crater remains constant with variations in spark duration a constant current density provides a more plausible explanation for the variation of crater diameter with spark duration than that the spark current density decreased with time as suggested by Somerville and Blevin.[13]

Figures 6 and 7. If the spark is considered as a point current source on the surface of the work piece and skin effect is neglected the current density J_r at any distance r from the point is

$$J_r = \frac{I}{2\pi r^2}$$

If a critical current density J_c is required for removal of a fragment at radius r the radius at which this current density exists will be

$$r = \sqrt{\frac{I}{2\pi J_c}}$$

and the radius of the crater should be

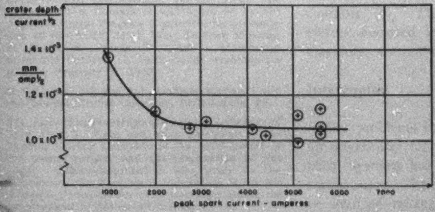


Figure 6. Effect of peak current of single oscillatory discharge of constant frequency on crater depth expressed as a function of ratio of crater depth to square root of current, showing dependence of crater depth on square root of current at higher peak currents

proportional to the square root of the spark current. Figure 6 is a plot of the ratio of crater depth to the square root of the current versus current and it is apparent that at higher currents crater depth is very nearly proportional to the square root of the current. A similar reasoning may be applied to show that volume of material removed, as in Figure 7, should be very nearly proportional to the three-halves power of the current. Deviations of the data from predicted forms at low currents may be explained as due to presence of surface imperfections of the same general magnitude as the craters and to a condition in which the craters are so small that the sparks cannot be considered point sources.

Numerous improvements in the speed and accuracy of spark cutting machines have been effected through redesign of these machines and their associated circuits in the light of the electrical field force theory.

References

1. Electro-Mechanical Method X Machines: Carbides, Hard Alloys, **M. F. Judkins, D. F. Dickey.** *Iron Age* (New York, N. Y.), volume 168, number 4, July 26, 1951.

2. Machining of Metals, **B. R. Lazarenko, N. I. Lazarenko.** *Stanki I. Instrument*, (U.S.S.R.), volume 17, number 12, 1946, pages 8–11, and volume 18, number 2, pages 4–8.

3. The Electrospark Method of Machining of Metals, **B. R. Lazarenko.** *Vestnik Machinostroeniya* (Moscow, U. S. S. R.), volume 27, number 1, 1947, pages 25–36.

4. Procédé et Machine pour le Travail de Métaux, **N. Lazarenko, B. Lazarenko.** Swiss Patent 257484, April 1, 1949.

5. Technology of Electric Spark Cutting of Metals, **D. T. Vasil'ev.** *Stanki I. Instrument*, (U.S.S.R.), volume 18, number 3, 1947, pages 1–8.

6. Electric Spark Machining, **I. Koncz.** *Gep*, (Hungary), volume 1, number 9–10, September-October 1949, pages 388–411.

7. Sharpening of Cemented Carbide Tools by the Electro-Erosion Spark Discharge Process, **B. N. Zolotykh.** *Stanki I. Instrument*, (U.S.S.R.), volume 18, number 3, 1947, pages 23–24.

8. Electronic Control of Electric Spark Machining Process, **D. T. Vasil'ev.** *Stanki I. Instrument*, (U.S.S.R.), volume 18, number 5, 1947, pages 1–5.

9. Calculation of the Electromagnetic System of an Electric Cutting Tool (Electric Spark Lathe), **V. N. Shchepetov.** *Elektrichestvro* (Moscow, U. S. S. R.), number 6, 20–30, June 1950.

10. Mechanism of Electro-Erosion of Metals, **S. L. Mandelshtam, S. M. Raiskii.** *Isvestiya Akademii Nauk SSSR* (Moscow, U. S. S. R.), Ser. Fiz, volume 13, number 5, 1949, pages 549–65.

11. The Electro-Fragmentation Process, **D. W. Rudorff.** *Engineers Digest*, (London, England), volume 10, number 9, September 1949, pages 306–08.

12. See, for instance, Modern Theory of Solids (book), **F. Seitz.** McGraw-Hill Book Company, New York, N. Y., 1940.

13. Current Densities in the Cathode Spots of Transient Arcs, **S. M. Somerville, W. R. Blevin.** *Physical Review* (New York, N. Y.), volume 76, number 9, 1949, page 982.

Discussion

I. Weber (Frankford Arsenal, Philadelphia, Pa.): In the paper Dr. Williams defined three processes of electromachining, and classified them as electrosparking, electroarcing, and electrolytic. The electrolytic process, as defined by Dr. Williams, consists of the work, usually cemented carbide, connected in a d-c circuit as the anode and held in contact with a rotating metal disk, which acts as the cathode. The work and the rotating disk are immersed in a solution of some sodium salt, usually sodium silicate in the form of commercial water glass. The process is presently being studied at the Frankford Arsenal, and it has been concluded that the process is electrolytic only in part. Identification of the process as an electrolytic process is misleading and therefore the term "electrolytic" is a misnomer.

When the process is applied as a method for sectioning hard metals and carbides, it has been concluded that the process is predominantly an arc cutting process. If a gap is maintained between the rotating disk and the work, the only phenomenon occurring is electrochemical, and is represented by the formation of a silicate film on the anode. There is no possibility for a self-initiated spark-over, since the voltages attained seldom exceed 20 volts. In actual practice a gap between the electrodes is constantly being opened and closed as a result of natural vibration of the apparatus. (The frequency of gap occurrence in the apparatus at the Frankford Arsenal has been measured to be 10 to 17 times a second.) At the moment the two electrodes separate, an arc is initiated, and is maintained until quenched either by too large a gap distance or by the sodium silicate solution. The arc is of relatively long duration, and because the arc has a high integral temperature, erosion of the anode is predominantly a thermal effect.

During one cycle of natural vibration, the following phenomena occur consecutively: arc erosion, electrochemical deposition of a silicate film on the anode, and direct short circuit. Arc erosion occurs at the moment the electrodes separate. When the arc is quenched, and before the electrodes make contact again to create a short circuit, a dielectric silicate coating is deposited on the anode. This coating is removed from the bottom of the cut by the rotating disk, thus exposing clean metal to further arc erosion. Sectioning proceeds thus in a very rapid sequential manner.

The application of the process to a grinding procedure has not yet been thoroughly studied at the Frankford Arsenal. However, there is no reason to believe that the process is purely electrolytic. It is very likely that the phenomena of electrodeposition, electroarcing, and electrosparking may all play significant parts to a greater or lesser extent.

Dr. Williams made the statement, probably obtained from Russian literature, that in an electroarc process erosion of the cathode is predominant. This may be true for arcing in air. The process being studied at the Frankford Arsenal exhibits a negligible amount of electroerosion of the rotating disc cathode. On the contrary, a build-up of anodic material on the sides of the disc has been observed. There is a noticable decrease in the diameter of the disk after extensive use, but the decrease is largely due to mechanical wear.

Everard M. Williams: Mr. Weber's comments relate not to the written paper but to additions in the author's oral presentation. Very little information has been available on the electrolytic process; in some Russian publications it has been confused with the electric spark process. Mr. Weber's disclosures are therefore of considerable interest.

With the current and developing shortage of industrial diamonds the importance of the various electric machining processes is increasing. Further fundamental research on these processes is needed.

A New Circuit Breaker for Aircraft Electric System

40-125 Amperes, 120 Volts A-C or D-C

B. S. BEALL, III
MEMBER AIEE

P. J. REIFSCHNEIDER
MEMBER AIEE

RECOGNITION of the need for better protection of aircraft electric systems is indicated by the increasing use of reverse current circuit breakers in d-c generator circuits. Improved protection, however, must be extended to the entire electrical system and not confined to the generating equipment. Protective equipment for low capacity circuits in 28-volt d-c systems has been available. There has been a lack of protective equipment, however, for similar and higher capacity circuits in both a-c and d-c 120-volt systems. A new circuit breaker with ratings up to 125 amperes, 115 volts, a-c or d-c, recently developed is described. The requirements for such a circuit breaker are:

1. Interrupting capacity, 5,000 amperes sea level to 50,000 feet.
2. Ultimate trip current, 138 per cent of rated current at 25 degrees centigrade.
3. Continuous current capacity, 115 per cent of rated current at 25 degrees centigrade.
4. Two hundred per cent current calibration, 10–100 seconds at 25 degrees centigrade.
5. Ambient temperature range, −55 degrees centigrade to 71 degrees centigrade.

A temperature compensated thermal element maintains the calibration of this circuit breaker within relatively narrow limits over a wide range of ambient temperatures. With minor changes, this circuit breaker can be arranged for manual operation and applied to 28-volt systems.

General Description

The circuit breaker, as shown in Figure 1, has an aluminum cover with mounting plates provided on one side. Power studs extend from an insulating base at one end of the circuit breaker. The control terminals of the closing and tripping solenoids extend through the cover at the opposite end. Exclusive of the power studs and mounting plates, the size is $7^{13}/_{32}$ by $5^{1}/_{4}$ by $1^{27}/_{32}$ inches. This remote operated circuit breaker weighs 4 pounds.

The circuit breaker with the cover removed is shown in Figure 2(A). Separate side frames support the mechanism and the thermal trip unit. The general arrangement of the parts is illustrated in Figure 2(B). At the lower left is the thermal element assembly, and above to the right are: the shunt trip (12); the closing solenoid (2); parts of the mechanism (3–8); and the interrupting chamber (9). The contacts are completely enclosed in this chamber, and therefore, are not visible here.

Interrupter

The problems involved in developing an interrupter for 120-volt aircraft electrical systems have been treated in an earlier paper.[1] Briefly, if some means can be used of raising the pressure of the atmosphere immediately around the contacts, the arc can be interrupted quickly even at very high altitudes in an interrupter of reasonable size.

A cutaway view of the interrupter and contacts is shown in Figure 3. An enclosure is formed for the contacts and interrupter by the cap and base molded of arc resistant phenolic asbestos compound. Silicone rubber gaskets (7) effectively seal the bottom of this enclosure. The moving contact consists of two parallel copper bridge members which are moved by the insulating operating rod (1). When closed, this contact completes the circuit between the stationary contacts; one on the bar attached to the thermal element, and the other on the connecting stud. Rapid opening to a total break distance of 3/4 inch is provided by the opening spring (2). The contact tip material, moving and stationary, is a form of silver tungsten carbide which is characterized by arc resistance, freedom from welding, and fairly low contact resistance.

Closely surrounding the movable and

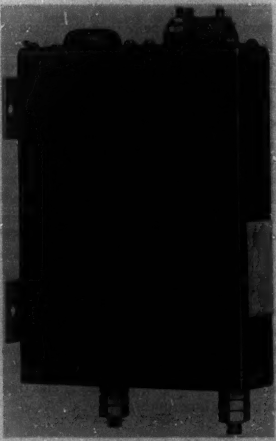


Figure 1. Aircraft circuit breaker, single pole, electrically operated, 120 volts, 400 cycle or direct current, side oblique view, cover on

---

Paper 52-131, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE South West District Meeting, St. Louis, Mo., April 15–17, 1952. Manuscript submitted October 11, 1950; made available for printing February 27, 1952.

B. S. BEALL, III, and P. J. REIFSCHNEIDER are with the General Electric Company, Philadelphia, Pa.

The authors acknowledge the helpful assistance of Miss Ruth E. Kern and other associates in the General Electric Company in the development of the circuit breaker and the preparation of this paper. Acknowledgment is also made to the members of the Equipment Laboratory of the Air Material Command for their assistance during the development of this circuit breaker.

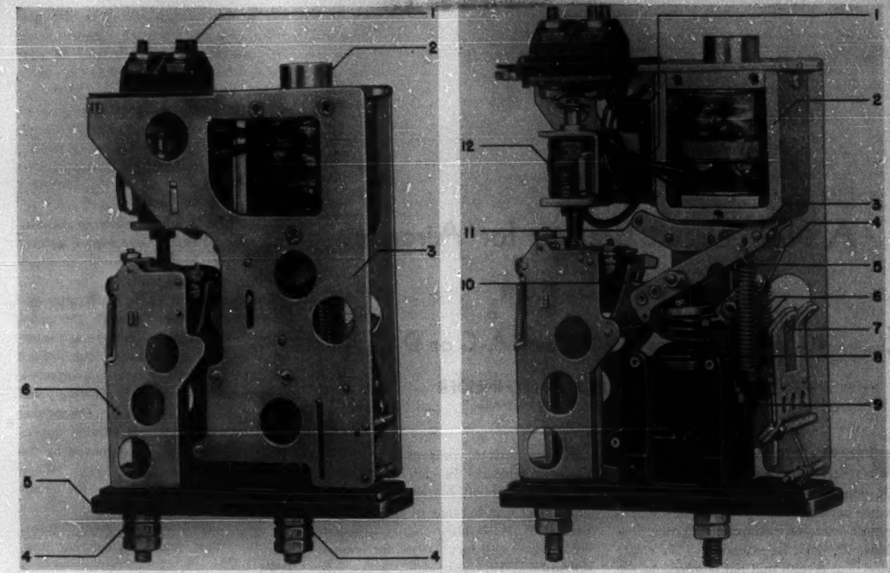


Figure 2. Aircraft circuit breaker, single pole, electrically operated

A (left). Side oblique view with cover removed

1. Terminal board
2. Closing solenoid
3. Mechanism side frame
4. Power studs
5. Compound base
6. Terminal element frame

B (right). Side oblique view with mechanism side frame removed

1. Auxiliary switches
2. Closing solenoid
3. Mechanism beam
4. Prop
5. Solenoid push rod
6. Contact pressure springs
7. Catch
8. Opening spring
9. Interrupting chamber
10. Latch
11. Auxiliary switch actuator
12. Shunt trip

stationary contact tips, and extending upward for slightly more than the length of the breaks, are tubes (6) of hard white vulcanized fiber. This material produces considerable gas when struck by an arc and does not deposit sufficient carbon to cause tracking. The blowout effect of the parallel arc streams causes the arcs, drawn between the contacts, to strike against the fiber at the higher current levels. At low current levels, the steel half tubes (8) around the fiber tubes help deflect the arcs against the fiber. A valve (4), molded on the insulating contact operating rod, seats against a valve seat (3) in the cap when the contacts are in the open position. The gas produced by the arc striking the fiber is confined in the interrupter by this valve and the gaskets (7) between the base and the cap. The resulting increase in pressure is effective in quickly interrupting the arcs.

Operating Mechanism

A floating beam mechanism has been developed to operate the contacts of this circuit breaker in a straight line. This mechanism is shown in Figure 2(B), with one of the supporting side frames removed. The contact operating rod is attached to the beam (3) at a point a little to the left of center. Under normal closing conditions, latch (10) blocks one end of the beam while the other is moved downward by the plunger of the closing solenoid (2). The connection between the pushrod of the solenoid and the beam is made through springs (6) which provide the required contact pressure. Prop (4) and catch (7) hold the mechanism closed until such time as latch (10) is released by the trip element, which allows the opening spring (8) to open the contacts quickly. In effect, the beam is latched at one end in closing, and is released at the opposite end when the circuit breaker is tripped. Although this mechanism is simple, it is trip-free. By trip-free is meant the contacts are able to open fully and quickly if the circuit breaker is tripped while the mechanism is held in the closed position. With this type of mechanism the flow of power cannot be maintained through the circuit breaker into a fault.

Closing Solenoid

To maintain the size and weight of the closing solenoid at a minimum, it has been designed with an operating winding of very short time rating. A second, higher resistance, winding is introduced into the circuit after the solenoid has closed the operating mechanism. The two windings in series may be continuously energized without overheating.

The circuit of the solenoid and short circuiting switch is shown schematically in Figure 7. When the solenoid plunger comes to the end of its stroke, the short-circuiting switch (1) in Figure 2(B), is opened and both windings are placed in series. As long as the solenoid is ener-

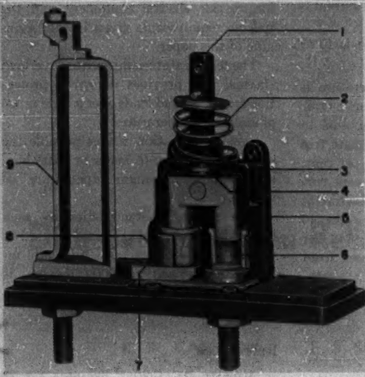


Figure 3 (left). Cutaway view of interrupting chamber

1. Contact operating rod
2. Opening spring
3. Valve seat
4. Valve
5. Bridge contacts
6. Fiber tubes
7. Gaskets
8. Steel half tubes
9. Expansion thermal element

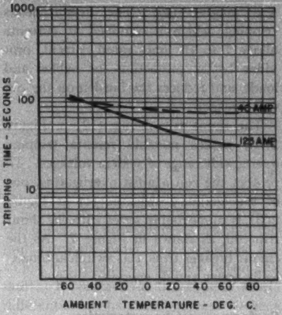


Figure 5 (right). Typical time—ambient temperature tipping characteristic at 200-per-cent current

gized, sufficient force is produced to hold the plunger in the closed position regardless of whether the mechanism remains closed or not. Only upon opening and reclosing the control switch will the solenoid attempt to reclose the mechanism.

This type of control scheme is often referred to as antipump, signifying that the closing solenoid, each time it is energized, will make but one attempt to close the circuit breaker.

Trip Devices

Overload tripping is accomplished in this circuit breaker by means of a temperature compensated linear expansion thermal element, shown in Figure 4.

The element proper (5) is made of material with a resistivity dependent upon the current rating of the circuit breaker. Current passing directly through the element provides the heat to raise its temperature to the tripping point. The small motion of the top of the element due to its linear expansion, is amplified approximately 20 times by an eccentric (3) and trip arm (10). This amplified motion is applied to the instantaneous trip armature (1), which, when rotated sufficiently, releases the latch. The force which the trip arm exerts is supplied by two tension springs. Compensation for changes in ambient temperature is provided by the expansion of the trip unit side frames; the temperature coefficient of expansion of these frames being similar to that of the thermal element. Changes in ambient temperatures produce approximately the same linear change in both the side frames and the thermal element. Since the trip arm is pivoted in the side frames, the resultant motion therefore, at the end of the trip arm, is negligible.

There is sufficient air space between the element and the side frames so that heat transfer between the element and the side frames during overload does not adversely affect the tripp'ng characteristic.

Combined with the thermal trip is a magnetic trip consisting of armature (1)

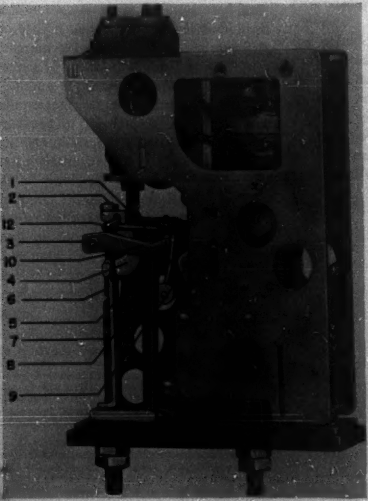
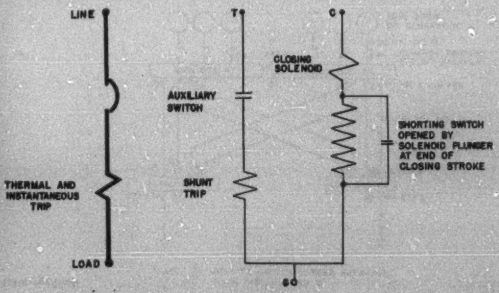


Figure 4 (left.) Thermal and instantaneous trip devices

1. Instantaneous trip armature
2. Calibrating screw
3. Eccentric
4. Instantaneous trip magnet
5. Expansion element
6. Operating spring
7. Side frame
8. Latch
9. Mechanism beam
10. Trip arm
11. Armature return spring
12. Calibrating screw
13. Compound base

Figure 6 (below). Circuit diagram for electrically operated aircraft circuit breaker


LINE
T
C
CLOSING SOLENOID
AUXILIARY SWITCH
SHORTING SWITCH OPENED BY SOLENOID PLUNGER AT END OF CLOSING STROKE
THERMAL AND INSTANTANEOUS TRIP
SHUNT TRIP
LOAD
G

and magnet (4). The instantaneous tripping, which is obtained at currents above 7 to 10 times rated current, provides better protection for cables and loads, and reduces the burden on the circuit breaker.

The shunt trip device (12) in Figure 2(B), makes possible remote tripping over a wide range of system voltages. To protect the coil against overheating, one of the auxiliary switches (1) in Figure 2(B), is connected in series with the shunt trip coil, as shown in Figure 7. As the circuit breaker trips, this switch automatically opens and de-energizes the coil.

Performance Characteristics

At 25 degrees centigrade the tripping characteristics of this circuit breaker are:

1. Tripping time at 115-per-cent current is in excess of 1 hour.
2. Tripping time at 138-per-cent current is less than 1 hour.
3. Tripping time at 200-per-cent is 10 to 100 seconds for all ratings.

The effect of ambient temperature is small at the ultimate trip current. Over the range of temperatures from −55 to +71 degrees centigrade the ultimate trip current is between 115 and 150 per cent.

The tripping time at 200 per cent current is affected by ambient temperature as shown by the typical curves for the 40 and 125 ampere ratings in Figure 6. Over a range of ambient temperatures from −55 to +71 degrees centigrade the tripping time is 10−200 seconds for all ratings from 40 to 125 amperes.

Conclusions

This newly developed aircraft circuit breaker meets the general requirements for aircraft electric equipment and the specific requirements given in the beginning of this paper.

The temperature compensated overload thermal trip provides improved protection for cables and loads over a wide range of ambient temperatures.

Remote operation makes possible the protection of remote loads or load busses with a minimum expenditure of weight.

The tripping characteristics are such that this circuit breaker may be used in either d-c or a-c systems. Adequate interrupting capacity, 5,000 amperes, is provided for all feeder and distribution circuit applications.

References

1. **Aircraft D-C Circuit Breakers for 28 and 120 Volts, B. S. Beall, III, P. J. Reifschneider.** *AIEE Transactions*, volume 69, part II, 1950, pages 1283–88

No Discussion

Distribution System Reliability of 28-Volt D-C Aircraft Electric Systems

K. W. CARLSON
ASSOCIATE MEMBER AIEE

E. S. SHERRARD
ASSOCIATE MEMBER AIEE

Synopsis: This paper analyzes the protection of a typical 28-volt d-c distribution system for a large multiengine aircraft. Methods of making a load study and fault study are illustrated. From the analysis of this system, conclusions are drawn regarding bus and feeder fault protection and arrangement. It is shown that by relatively minor modifications in bus arrangement a considerable increase in the degree of reliability can be obtained.

PRESENT-DAY aircraft are dependent for best operation on the reliable, continuous service of their electric systems. The design, installation, and maintenance of these systems must be kept within certain standards to achieve the desired results. The importance of protecting the system by an integrated fault protection design and a well-disciplined program of installation and maintenance cannot be overemphasized.

Figure 1 shows diagramatically the equipment and circuits of the system analyzed in this paper. In general, it consists of six generators, two batteries, and a centralized bus system and is protected primarily by the use of current limiters. Detailed descriptions of systems similar to the typical one analyzed herein may be found in other papers.[1,2]

In preparing or checking the fault protection of an electric system it is necessary to consider at least three types of faults.

1. Continuous short circuits.
2. Continuous arcs with arc voltage 14 to 20 volts.[3]
3. Intermittent arcs.

Continuous short circuits or solid faults may occur as welded faults or may result from mechanical failure of a bus structure or the failure of some other protective device. This type of fault presents a severe problem in systems using self-excited shunt machines because of the low-generator current available for clearing the fault at the low fault resistances involved. This steady-state current ranges from about 200 to 500 amperes per 300-ampere generator.

Continuous arcs, if maintained, are cleared more easily because the voltage involved (14 to 20 volts) moves the operating point on the generator volt-ampere curve toward the knee of the curve and the fault currents are in the range of 500 to 1,000 amperes per 300-ampere generator. It is possible that a fault of this type will burn clear before the limiter is able to clear the circuit.

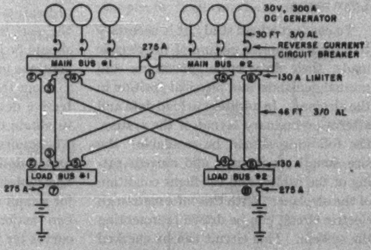
Intermittent arcs are not analyzed in this paper. In previous investigations the fault currents under such conditions are so low that limiters could not detect the fault. To protect for this type of fault, differential or balanced current devices may be necessary. Differential protection now is being applied to some important circuits, such as the generator circuit, but has not been applied generally to the distribution circuits.

Criteria for System Protection and Operation

The desired characteristics of a fault protection scheme should include:

1. Faults should be cleared as quickly as possible and should be interrupted before damage is done to system components.
2. Protective devices should operate in such a manner as to isolate only the faulted section, thereby retaining the maximum amount of available feeder and generator capacity.
3. A well-protected system should have adequate backup protection in the event of failure of any one of the protective devices.
4. The system should operate without false or nuisance trips.
5. Satisfactory performance as stated in these items should be maintained even with the prior loss of some capacity and/or the loss of certain feeders.

Fast operation of protection will minimize the fire hazard concurrent with faults and will give greater system reliability. Efforts have been made in the past and will be made in the future to



NOTE: ① DENOTES LIMITER NUMBER REFERRED TO IN TABLES I AND II

Figure 1. System diagram

Paper 52-132, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE South West District Meeting, St. Louis, Mo., April 15-17, 1952. Manuscript submitted October 19, 1950; made available for printing February 27, 1952.

K. W. CARLSON is with the General Electric Company, Schenectady, N. Y. E. S. SHERRARD was with the General Electric Company, Schenectady, N. Y., at the time this paper was written, and is now with the Naval Research Laboratory, Washington, D. C.

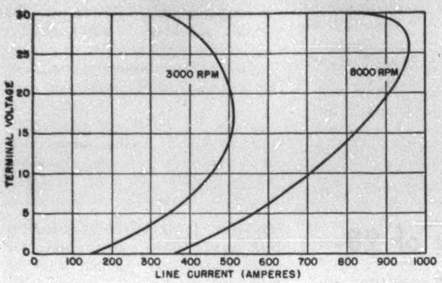
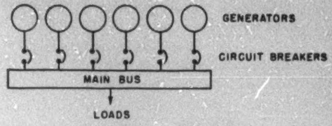
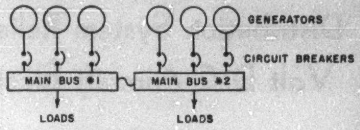


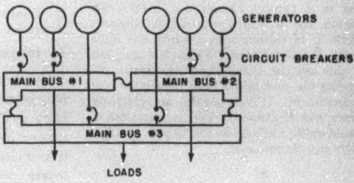
Figure 2 (above). Volt-ampere curve for a 30-volt 300-ampere d-c generator



SINGLE BUS



TWO-SECTION BUS



RING BUS

Figure 3 (right). Main bus arrangements

lessen the operating time and improve the selectivity of protective devices.

Backup protection may be considered as insurance to protect against faulty devices elsewhere in the system.

The criteria of having the system operate under conditions of reduced capacity and inoperative feeders is an important safety factor consideration. Aircraft operators, in their operating procedures, may allow take-off with one of six generators inoperative and it is possible that on the same flight another may fail. Protection for this condition must be provided for, just as it must be where one or more feeders may not be available for use because of a previously cleared fault.

Method of Study

RATINGS OF PROTECTIVE DEVICES

The sizes of the fault protective devices are determined largely by the magnitude of load carried by them during normal, abnormal, and emergency operation. Therefore, it is usually necessary to make a load study of the system before fault protection elements can be chosen. Then from the results of the fault study the ratings of the protective elements are adjusted to give the maximum co-ordination.

LOAD STUDY

To conduct a load study, it is necessary to have a complete picture of the electric loads, their duty cycle, maximum and normal demands, and physical position in the airplane. In addition to load data and after a preliminary layout of the system, the following should be available: the size, length, resistance, and current rating of the cables, and the flight condition of the airplane. With this information an electric circuit can be drawn representing the system. This circuit can be checked for voltage regulation and for current-carrying capacity in each branch, thus determining the limiter sizes. The system also should be checked for specified abnormal or emergency operation. Limiters should be selected to carry the maximum load currents found under emergency conditions, but should be as low a value as possible for fast clearing under faults.

FAULT STUDY

The effect of abnormal conditions on the system are checked by a fault study. In this study, faults should be placed on the system at various positions and the resulting currents in each branch of the circuit noted. Then the limiter time-current curves are consulted to determine the co-ordination of the system. That is, only the part of the system which is faulted should be isolated; the remainder should continue to supply the load.

From this study, the sizes of the protective devices may have to be adjusted to obtain the best co-ordination possible. The results may also indicate the way to a better design or arrangement of busses and cables. The same data are necessary for a fault study as for a load study with two major additions, the time-current curves for the protective devices and the volt-ampere curves which serve as the representation of the generators. The addition to steady-state fault currents from the initial transient can be neglected if the clearing times of the protective devices are three to four times greater than the time constants of the generators.

Figure 2 illustrates the volt-ampere curve for a 30-volt 300-ampere d-c generator with voltage regulator control. Since it is nonlinear, it cannot be represented by any simple mathematical expression. The curves must be obtained by test and accuracy must be maintained, especially for the lower part of the curve, since the operating point during faults will fall on this part of the curve.

The calculations may be done either by trial and error hand solution or by use of the d-c network analyzer.[4] In either method it is advisable for simplicity to use a per-unit system. For use of the d-c network analyzer the normal source voltage is taken as 1.0-per-unit voltage and the normal rating of a generator as 1.0-per-unit current. The per-unit resistance may be calculated from the per-unit voltage and current. In self-excited systems, since the voltage of the generator falls to a low value at short circuit, the highest source voltage is the battery and is represented by a 1.0-per-unit voltage behind its internal impedance. The circuit resistances are changed to per-unit values and the volt-ampere curves may be combined and also replotted in per-unit values. The analyzer ground connection is placed at the point of fault and the resistance representing the generators adjusted to read the proper values from the curves. Currents are read in each line, changed to actual amperes, and checked against the time-current curves to establish which limiter or limiters will melt first. If the other limiters in the circuit had only a small amount of fault current heating them, a second calculation must be made with the first limiter open to establish which limiter opens next. Generally a fault will be cleared by the opening of two limiters or sets of limiters, and thus two sets of calculations per fault are necessary. The total clearing time under these conditions is the sum of the melting times of the two calculations.

If the results warrant changes, the sizes of the protective devices should be altered to obtain the best co-ordination possible for the design. The results also may suggest changes in the arrangement of busses or cable runs which may greatly improve co-ordination and reliability.

Discussion

Bus Configuration

Figure 3 illustrates three types of main busses which we shall refer to as single, 2-section or split, and ring type. The first of these, the single bus, is the simplest. However, its use makes the system extremely vulnerable to a solid main bus fault. Such a fault would cause a loss of the whole system.

The split bus illustrated has the great advantage that a solid main bus fault results in loss of only half of the main bus load. Once the faulted bus is isolated, all loads except those connected to the faulted bus may be carried by the other main bus section. It is important to note that the vital instrument and radio loads ordinarily supplied through the load busses would not be lost. The employment of this bus rather than the single bus involves, in most cases, little or no increase in weight or space.

The use of a ring bus further decreases system vulnerability to a main bus fault. On the other hand, it increases system complexity with resultant design problems relating to the number, size, and run of cables, and the obtaining of equal division of the main bus load for all conditions of flight and reduced generation operation. If a ring bus is to be used, a check always must be made to determine that the bus tie overcurrent devices will not be opened or damaged by bus tie currents due to unbalanced bus loads, a reduced number of generators on the bus, or a combination of both.

A common danger in all these systems is that a faulted bus or bus section will not be isolated from the generators connected to it, possibly resulting in a fire and damage or loss of equipment. The short-circuit current of a generator, as shown in Figure 7, may be too small to operate overcurrent protection. Additional devices such as undervoltage or polarized relays connected between bus sections may be used to secure this isolation. A scheme using polarized relays is drawn in Figure 4 for a split-bus system. After the limiters open to isolate bus Number 1 from the rest of the distribution system, the flow of current through the polarized relay *A* opens the reverse current circuit breakers of the generators feeding bus Number 1 and completes its isolation. A polarized relay appears preferable to an undervoltage relay since it gives faster clearing and there is no problem in co-ordinating relay time delay with system transients or time-current characteristics of overcurrent protective devices.

Use of Battery to Aid in Fault Clearing

For a fault on main bus Number 2 of Figure 1, the bus tie limiter opens first. Fault current then flows through the feeders between main busses and load busses.

Without batteries on the load busses, the same current would flow through the limiters protecting main bus Number 2 and those protecting main bus Number 1. Both sets of limiters would be damaged and either or both sets could clear. With the batteries on the load busses, the feeders to main bus Number 2 from the load busses carry more current than do the feeders from main bus number 1 to the load busses. This additional current assures co-ordination between the limiters protecting main bus Number 2 and main bus Number 1 and results in a minimum loss of system load and generator capacity.

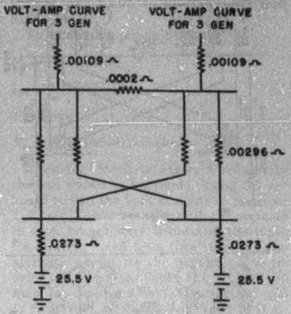


Figure 5. System equivalent circuit

Relation Between Fault Protection and Cable Protection

In the distribution system studied overcurrent protective devices were chosen in accordance with the methods advocated in this paper. Protective devices which are chosen on the basis of load and fault studies are usually smaller than necessary to protect the cable. This is particularly true if the cable size is restricted by having to maintain the proper voltage drop from generator to load bus. If the protective devices were chosen only on cable size, the cable itself could be protected but the system would not necessarily be fully protected from faults. For example, limiters chosen to protect the feeder cables between main and load busses could be so large that they would fail to open for faults on the main bus. Thus

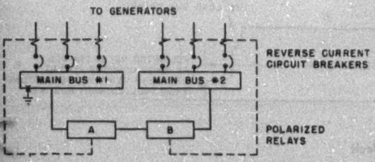


Figure 4 (left). Polarized relay scheme for removal of generators from faulted bus

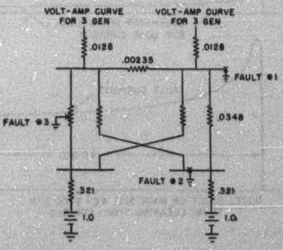


Figure 6 (right). System per-unit circuit

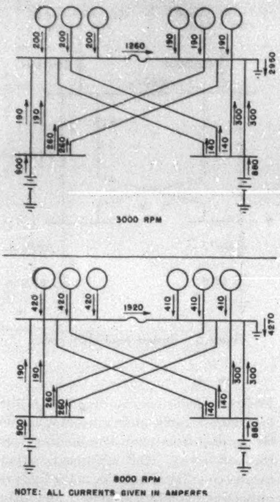


Figure 7. Fault currents for fault on main bus number 2

the whole system would be lost and the safety of the airplane jeopardized.

A desired result, as stated in the criteria for fault protection, is to obtain both protection of the system and all elements of the system. A fault study will determine whether or not the system has the proper co-ordination of protective devices.

CABLE FAULT PROTECTION BY USE OF MULTIPLE CABLES

The use of multiple cables with limiters at their extremities was found to be an effective means of isolating faulted cables. At the same time it assures a supply of power to the load even in the case of failure of one of the cables.

Table I. Summary of Results

Operating Conditions: Night Take-off, 1,750-Ampere Total Load, 700-Ampere Load on Bus Number 2, 600-Ampere Load on Bus Number 1, 225-Ampere Load on Each Load Bus

Case Number	Number of Generators		Amperes in Feeders			Bus Tie, Amperes	Generator Amperes
	Bus Number 1	Bus Number 2	L_2	L_3 and L_5	L_4 and L_6		
Continuous loads	3	3	56.3	56.3	56.3	50	292
Line L_2 out	3	3	0	75	56.3	88	292
One generator removed	3	2	56.3	56.3	56.3	225	350
One generator removed, 5-second transient motor load on bus number 2	3	2	56.3	56.3	56.3	465	430

SYSTEM RELIABILITY

One method of increasing the reliability of the system studied would be to protect busses and cables by differential or balanced current devices. This method would give faster clearing times and would protect for smaller fault currents than does the present protection. On the other hand its use would mean a large number of devices and increased weight, space, and cost.

Since most systems in use today and systems for the near future rely principally on inverse time overcurrent protective devices for protecting distribution systems, the system chosen for study was of this type. The numerical results obtained apply to this particular system. But the majority of the conclusions are general and apply to aircraft systems which are protected for the most part by inverse time overcurrent protective devices. The study illustrates that evaluation of system reliability for large aircraft requires thorough load and fault studies for the system. Results of such studies point out system deficiencies, suggest ways of remedying these deficiencies, and enable positive conclusions to be drawn regarding system performance and reliability.

Conclusions

The following conclusions may be drawn:

1. A 2-section bus appears to be the best main bus arrangement for ease of fault clearing and general reliability of the system.
2. Symmetrical arrangement of feeders to a load bus from both sections of the main bus results in a better co-ordinated system.
3. For a split-bus system, which employs inverse time overcurrent protective devices, battery location at the load bus permits better co-ordination of those protective devices for main bus faults.
4. Protective devices may be best selected for system protection by using a fault study to determine whether or not these devices are co-ordinated for all fault conditions.
5. System reliability can be evaluated by load and fault studies of the type illustrated.

Appendix I. Sample Calculation for Load Study

For the purposes of a load study number the lines between main and load busses $L_2, L_3, \ldots, L_6$ corresponding to limiters numbers 2, 3, 4, and so forth, as in Figure 1. The loads under night take-off are considered as maximum load for the study.

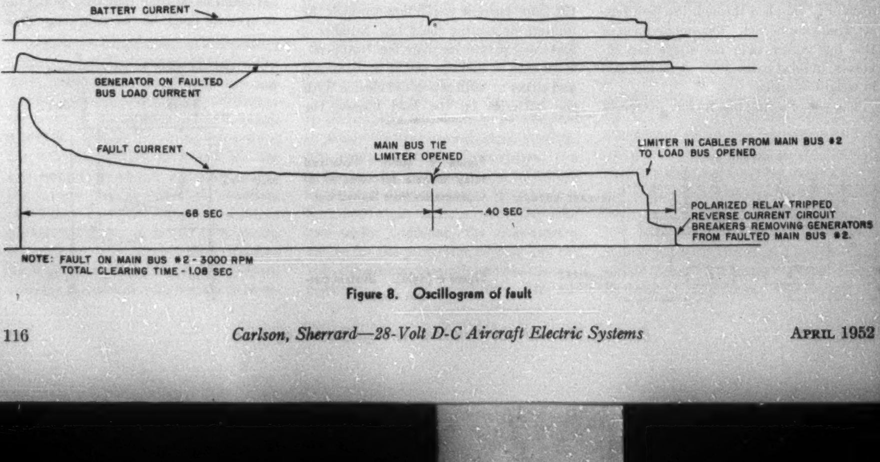


Figure 8. Oscillogram of fault

Case I

Total load, 1,750 amperes
Main bus number 1 load, 600 amperes
Main bus number 2 load, 700 amperes
Load bus number 1 and number 2 load, 225 amperes each

I for lines 2, 3, 4, 5, 6 = 450/8 = 56.3 amperes

I generator = 1,750/6 = 292 amperes

Summing all currents at main bus number 2

I for bus tie = 700 + 225 − 3 × 292 = 50 amperes

Case II

The conditions are the same as for case I except that line 2 is open. The current in lines 4 and 6 are the same as case I but the current in lines 3 and 5 is

I for $L_{3,5}$ = 225/3 = 75 amperes

I for bus tie = 700 + (2 × 75) + (2 × 56.3) − 875 = 88 amperes

Case III

With two generators on bus number 2 and three on bus number 1 the currents are as follows

I generator = 1,750/5 = 350 amperes

I for bus tie = 700 + 225 − 2(350) = 225 amperes

Case IV

The load in case 1 is the continuous value. One component of it is a compressor motor supplied by main bus number 2. Its continuous current is 200 amperes and its transient current is 400 amperes flowing for 5 seconds. The generator current during this period for one generator removed from bus number 2 is

1 generator = (1,300 + 400 + 450)/5 = 430 amperes

1 for bus tie = 1,100 + 225 − 860 = 465 amperes

These results are summarized in Table I. Referring to cable data it is found that these current magnitudes do not exceed the cable rating. The currents permit using a 275-ampere (continuous rating) bus tie limiter and a 130-ampere (continuous rating) limiter for lines L_2, L_3, L_4, L_5, and L_6.

Table II. Fault Study Results

Fault	Rpm	Limiters Which Melt	Total Clearing Time, Seconds
1	3,000	1, 5, 6	1.08
	8,000	1, 5, 6	0.54
2	3,000	4, 6, 8	1.50
	8,000	4, 6, 8	1.50
3	3,000	2	0.05
	8,000	2	0.02

Note: Refer to Figure 1 for limiter numbers.

Appendix II. Sample Calculations for Fault Study

The electric circuit resistances shown in Figure 5 can be calculated and reduced to one line from the information in Figure 1. The volt-ampere curves referred to are shown in Figure 2. An average value of voltage and internal impedance for the battery was found to be 25.5 volts and 0.024 ohm.

To establish a per-unit system

V_{pu} = 25.5 volts

I_{pu} = 300 amperes

Therefore

R_{pu} = 25.5/300 = 0.085 ohm

Then the circuit in Figure 5 reduces to the one shown in Figure 6. Using the d-c board to find the fault currents at positions 1, 2, and 3 shown on Figure 6, the currents in each branch were found and the times for clearing noted from time current characteristic curves for the limiters. The results are found in Table II. A fault on main bus number 2 for 3,000 and 8,000 rpm is illustrated in Figure 7. Figure 8 is an oscillogram of a main bus number 2 fault in a system using a polarized relay scheme for removing the generator from the faulted bus.

References

1. Present D-C Aircraft Electric-Supply Systems, H. C. Anderson, S. B. Crary, N. R. Schultz. *Electrical Engineering* (*AIEE Transactions*), volume 63, June 1944, pages 265–272.

2. D-C Power Systems for Aircraft, H. J. Finison, R. H. Kaufman. *General Electric Review* (Schenectady, N. Y.), September, October, November, December 1945, pages 22, 35, 59, 56; March, May, June, August 1946, pages 51, 46, 46, 41.

3. Electric-Circuit Burning-Clear and Damage Phenomena on Aircraft Structures, C. M. Foust, J. G. Hutton. *Electrical Engineering* (*AIEE Transactions*), volume 63, April 1944, pages 198–204.

4. Analyzers, C. Concordia, H. A. Peterson. *General Electric Review* (Schenectady, N. Y.), volume 48, number 9, September 1945, pages 29–37.

No Discussion

Weight Analysis of Aircraft Actuators

HOWARD M. GEYER
NONMEMBER

R. C. TRESEDER
ASSOCIATE MEMBER AIEE

THE work that must be expended to actuate adjustable flight surfaces and to operate the landing gear, bomb doors, and other movable structures on aircraft often exceeds pilot capabilities. At the same time, the increase in physical size of aircraft in recent years has all but ruled out mechanical transmission systems. To provide for power amplification and to eliminate mechanical links, the aircraft designer uses a multitude of remotely controlled actuating devices, known more simply to the aircraft industry as actuators.

As an externally supplied power source, these actuators use either hydraulic, electric, or pnuematic energy and may in some cases use a combination of the two, or even all three.

The way in which the external energy or working fluid is controlled differentiates between the two general types of actuators. The first is known as the open-loop type and is illustrated in Figure 1(A). The pilot or flight engineer, as the case may be, operates the controlling valve. As long as the control valve is opened, or switch is kept closed, the actuator continues to move. The movement of the actuator is either terminated by the pilot's realization of the device's objective, that is, proper trim angle, and so forth, or by limit switches or stops in the controlled assembly. In a sense, this actuator has some of the elements of a closed-loop system. The second general type of actuator utilizes control position to designate a desired position of the actuator and the actuator itself, by means of the closed loop, initiates the signal that causes the control to stop the flow of the power fluid to the actuator.

For purpose of this paper, the discussion will be limited to actuators controlled by an open-loop control system. This was deemed advisable as in the closed-loop type, the control itself becomes a major item of weight and complexity.

The work of an actuator can vary from a few hundredths of a horse-power, in the case of a cowl flap actuator, to 10 or 20 horsepower in the case of a large landing gear actuator or a flight surface actuator for a high-speed airplane.

In comparing the three types of power fluids, electric, hydraulic, and pneumatic, and the type of generation, transmission, and actuation that may be used, a multitude of variables at once becomes apparent. To simplify the presentation, one very controversial point of comparison will be avoided entirely. That point is reliability. Improvements are being made so rapidly in each of the systems that statements made today on reliability would be out of date and possibly incorrect in a few months. Cost comparison, while it may be a deciding one, will also be avoided since there are many variables that determine costs, such as production rates, installation expense, use of standardized units, and so forth. The comparison will be made on the basis of weight, size, and over-all efficiency.

The plan will be to study first the generation of the power fluid; second, the storage of the power fluid; third, the transmission of the power; and fourth, the utilization of the power fluid by the actuator.

Generation

Electric generators for aircraft have become fairly well standardized over the past few years and it was possible to obtain from the respective manufacturers rather consistent data on both weight and efficiency. Figure 2 is a plot of specific weight, that is pounds per horsepower, versus horsepower for both d-c and a-c generators. This data can be plotted logarithmically as almost a straight line. The trend of reduced specific weight as horsepower is increased is apparent. This may be somewhat misleading, as the gains made in the larger machines are not due to size alone. The larger generators have been built in later years and reflect improvements in insulation and the effects of higher operating temperatures, blast cooling, and higher drive speeds. Perhaps the curve should show a distinct leveling-off at the higher powers.

There is a definite reduction in weight of the a-c generator over the d-c version. This is especially true in the larger sizes. The specific weight of large sizes appears to level off at 2.0 pounds per horsepower. The short term overload capacity of an electric generator for a period of as much as 5 minutes is approximately 125 per cent, and this allowance should be made in selecting a generator for most actuator requirements. The weight and the constant speed drive of course will penalize the a-c generator. The requirements of a constant speed drive will vary between piston engines and gas turbines. The latter will permit a smaller variation between normal and maximum speed. The generator over-all efficiency shows an improvement with size in both the a-c and d-c types with an average value of 73 per cent for the smaller sizes and a mean of 83 per cent for the larger sizes.

Also plotted on Figure 2 are the specific weights of 1,500- and 3,000-pound-per-square-inch hydraulic pumps of the engine-driven types. Again there is much the same trend as the electric generators. As there is a wide variation in the types of hydraulic pumps and in the associated valving required to maintain the desired flow, it is more difficult to establish the

---

Paper 52-133, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE South West District Meeting, St. Louis, Mo., April 15–17, 1952. Manuscript submitted September 10, 1950; made available for printing March 14, 1952.

HOWARD M. GEYER and R. C. TRESEDER are with the General Motors Corporation, Dayton, Ohio.

The authors gratefully acknowledge the assistance of J. N. Tootle and F. L. Ruhl, both of the Aeroproducts Division of General Motors Corporation.

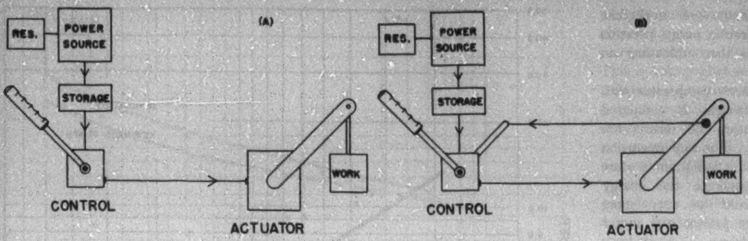


Figure 1. Open- and closed-loop systems

mean curve. It was decided to neglect the valving weight, as the voltage regulator weight also had been neglected in the case of the electric generator. As is to be expected, the higher pressure pumps show a lower specific weight because the delivered horsepower of a constant displacement pump will vary directly with pressure. In a constant pressure system, the hydraulic pump will have no overload capacity as far as output horsepower is concerned. The specific weight of the 1,500-pound-per-square-inch pump has an asymptotic value of 1.0, while the value for the 3,000 pound-per-square-inch pump is approximately 0.5.

The over-all efficiency (volumetric plus mechanical) of a hydraulic pump at rated load is not too much different from an electric generator. Values of 75 per cent for small sizes to values as high as 90 per cent for larger sizes may be obtained. The part load efficiency like that of the electric generator is lower than the full load value.

There is less available data concerning high-pressure pneumatic compressors for aircraft use than there is on generators and pumps. The high compression ratio needed, especially for high altitude operation, requires the use of several stages. For that reason, high pressure compressors are not built in as large sizes as are either the hydraulic or electric generators. The specific weight of this third type of generator was outside the range of Figure 2. The specific weight of one make of compressor varied from 186.5 pounds per horsepower in the 0.016-horsepower size to 40.1 pounds per horsepower in the 0.523-horsepower size. The pneumatic compressor also shows a lower efficiency than either of the other two.

Power Storage

Aircraft actuators are used intermittently, and the integrated power demand of a series of actuators is so small that power storage appears, at first glance, to be attractive. Electric energy for aircraft is stored in much the same way as it has been in automotive use for the past 30 years. The lead storage battery for aircraft use is only a fraction lighter than the auto battery for a given energy storage, and batteries built by different manufacturers show the effects of standardization. Table I shows the amount of energy that may be obtained from an aircraft battery at different discharge rates.

The ability of a storage battery to deliver its stored energy is dependent upon the rate at which it is to be delivered. Increasing the rate drastically reduces the amount of energy that can be removed from the battery at one time. For the short time periods, however, the battery will recover and, except for the greater I^2R losses, will deliver approximately the same total energy whether that energy is taken in short spurts at high current or as a continuous drain at a low current.

The weights of both hydraulic and pneumatic accumulators are shown in Figure 3. Both types of accumulators show a reduction in specific weight in the larger sizes. The method by which both air and oil are stored under pressure is well known and will not be discussed here. The weight values used in plotting Figure 3 were supplied by manufacturers of aircraft accumulators and show a definite similarity. The pneumatic accumulators have lower specific weight as can be expected since all the volume is usable. The specific weight of both types of accumulators includes the weight of the pressure working fluid plus the weight of the compressed gas of the hydraulic accumulator.

It is interesting to note that the electric storage battery is far superior as a storage medium to either of the other two when discharged at either its 5-minute or its 5-hour rating.

To assist in comparing the relative weight of either a generator or storage medium in providing power for a given actuator, Figure 4 was plotted. The battery line does not cross the generator line for any period of discharge. For a load with a time duration of 20 seconds, a battery weighs three times as much as a generator with equal capacity. For this reason, a storage battery is used only to provide power when the primary source of power is inoperative and to offer some voltage stabilization during fluctuating loads.

A hydraulic pump is also lighter than an accumulator except for a demand of less than 2 seconds. This indicates that the accumulator is used not for the storage of energy but to absorb shock or surge in the system and to simplify the problem of unloading the hydraulic pump when

Table I. Comparison of Storage Battery Performance

	Automotive Battery	Transport Battery*	Jet Aircraft Battery
Five hours, 80 degrees Fahrenheit, ampere-hours	76.5	88	68.5
Weight per cell, pounds	13.6	13.0	12.3
Five hours, ampere-hours per pound	5.6	6.7	5.5
Five minutes, 80 degrees Fahrenheit, amperes	324	357	442
Five minutes, watt-minute per cell	2,620	2,760	3,540
Five minutes, watt-minute per cell per pound	192	212	288
Space per cell, cubic inches	187	109 (147)	165
Five hours, ampere-hours per cubic inch	0.4	0.8 (0.7)	0.4
Five minutes, watt-minute per cell per cubic inch	14	25.3 (18.7)	21.5

* Figures in parentheses are for batteries with nonspill feature.

the system demands are negligible. Unloading an hydraulic pump presents more complications than unloading an electric generator.

When the pneumatic compressor with its high specific weight is compared with its accumulator, the results are strikingly different. For any operation less than 4 minutes, it is lighter to store rather than to generate the energy directly. This would be even more pertinent if the compressor was motor rather than engine driven.

From Figure 4, the following conclusions can be drawn:

1. Electric and hydraulic systems will use energy storage only for extremely short duration operation and for surge protection.
2. The pneumatic system will use large storage capacity and the minimum size of compressor possible.
3. Power for actuator on when the prime source of power is shut down should be provided by a battery for long term demands, that is, over 20 seconds, and by either a hydraulic or pneumatic accumulator for short duration demands. For repetitive short term demands, it is desirable to use an electrically driven pump to charge an accumulator rather than either a large accumulator or a large battery with a high discharge rate.

Transmission

The electric system in aircraft uses a grounded return for d-c transmission and a 3-wire grounded neutral for 3-phase a-c transmission. Conductor size is determined by current density as well as voltage drop, and rigid specifications define the allowable limits for each. In military aircraft, conduit is seldom used. The computation of wire weight for any amount of power to be transmitted is very simple. Aluminum is used for the larger sizes of cables to save weight. The use of high voltage, either alternating or direct current, is imperative in large aircraft for anything but nominal size actuators.

Aluminum tubing is used for both hydraulic and pneumatic transmission in those areas not subject to temperatures above 300 degrees Fahrenheit. For higher temperatures, stainless steel tubing may be used.

Figure 5 is a plot of the weight necessary to transmit power 75 feet by the three mediums. The electric system uses a grounded return and a single power lead plus two control leads. It is based on a maximum voltage drop of two volts on a 24-volt system and on current ratings of the army-navy specification *AN-W-14a*.

The hydraulic system has both a

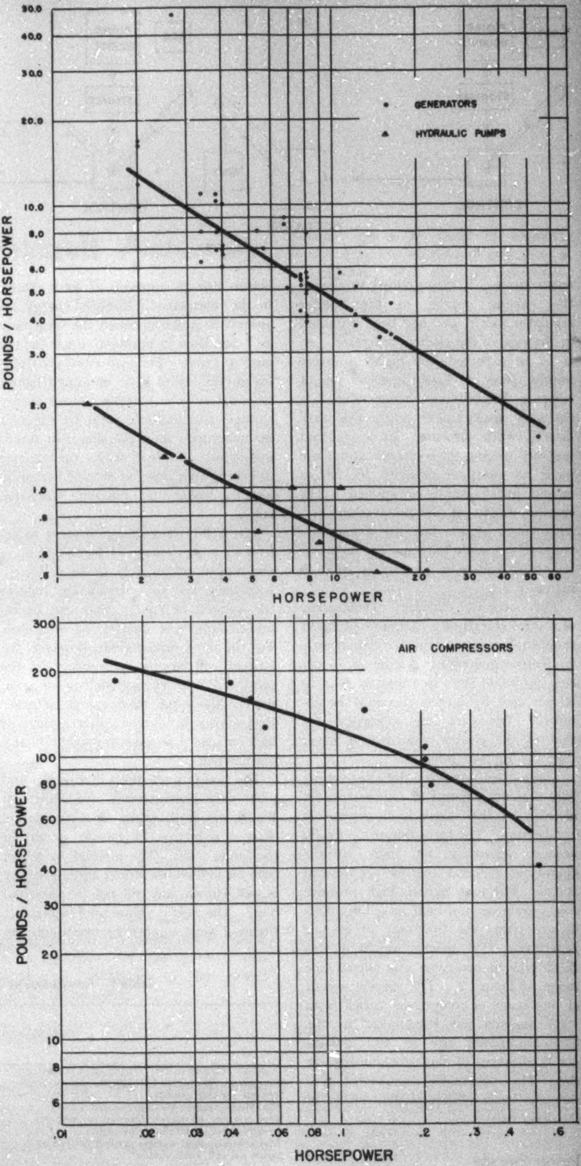


Figure 2. Specific weight of generators

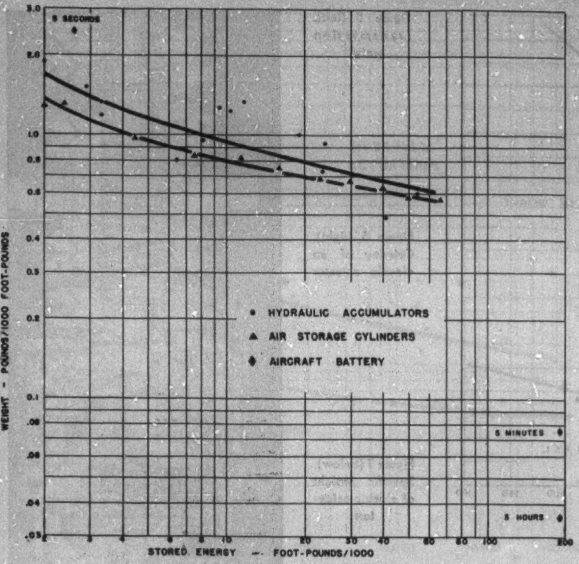


Figure 3. Specific weight of storage media

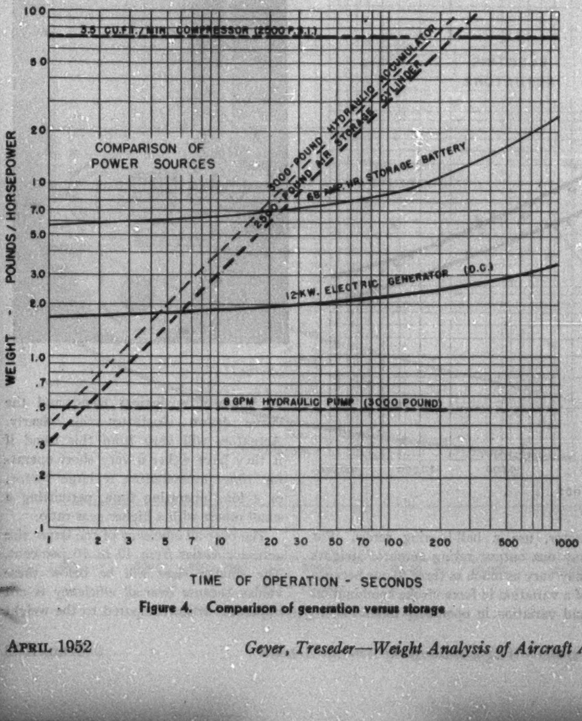


Figure 4. Comparison of generation versus storage

pressure and sump line. Tube size is chosen to give a transmission efficiency of 90 per cent at −20 degrees Fahrenheit. This is a somewhat severe basis for computation. An aircraft designer will compromise somewhat on transmission efficiency to obtain a minimum over-all system weight. The specific transmission weight of energy at 3,000 pounds per square inch is lighter than at 1,500 pounds per square inch for the same tubing stress and so is used in this comparison.

The pneumatic system uses a single transmission line. This is made possible by exhausting at the actuator directly to the atmosphere. An operating pressure of 1,500 pounds per square inch was chosen and the transmission efficiency was computed at 90 per cent. Temperature has a negligible effect.

The weight of both hydraulic and pneumatic transmission systems includes the weight of the working fluid.

Utilization

The means by which power transmitted to the actuator is transferred into useful work by the actuator is dictated by the mechanical design of the structure to be moved. The output motion may be either linear or rotary and the stroke-force relationship will vary over wide limits. In the larger sizes, however, linear actuators fit into the structure more conveniently. In this paper, linear actuators only will be considered.

Figure 6 is a section view of a ball-bearing type jack screw actuator of the electrical type. The motor, gear reduction, and screw are coaxial and the complete assembly is installed much the same as a hydraulic cylinder. The motor is totally enclosed and is rated for short-time intermittent duty. The motor case is part of the structure. Because of the high efficiency of the ball-bearing jack screw, even in the reverse direction, and the high efficiency of the gear reduction assembly, the motor is provided with a magnetically released brake to prevent the unit slipping under load. The motor diameter-length ratio is kept low to reduce starting time and to minimize the kinetic energy to be absorbed in the brake.

The size of the ball-bearing jack screw is dictated by the brinelling capacity of the nut as well as by column strength of the screw. If the actuator is operating purely in tension, the screw can be smaller in diameter.

In Figure 7 is plotted the specific weight of a series of approximately 50

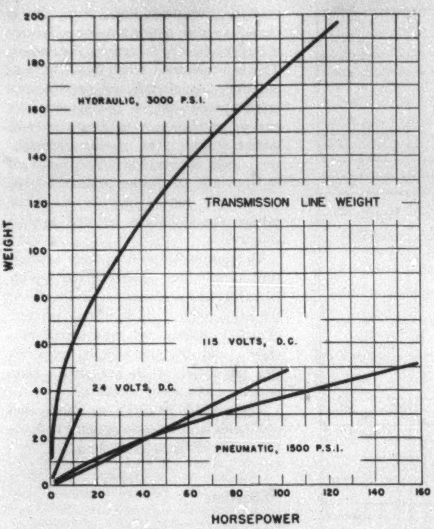


Figure 5 (left). Transmission weight

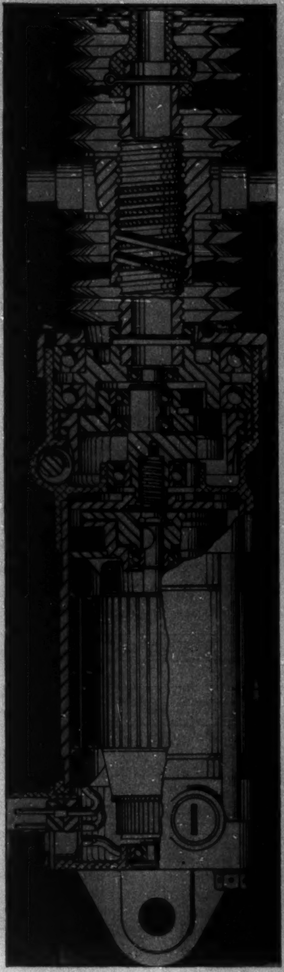


Figure 6 (right). Cutaway of an electric actuator

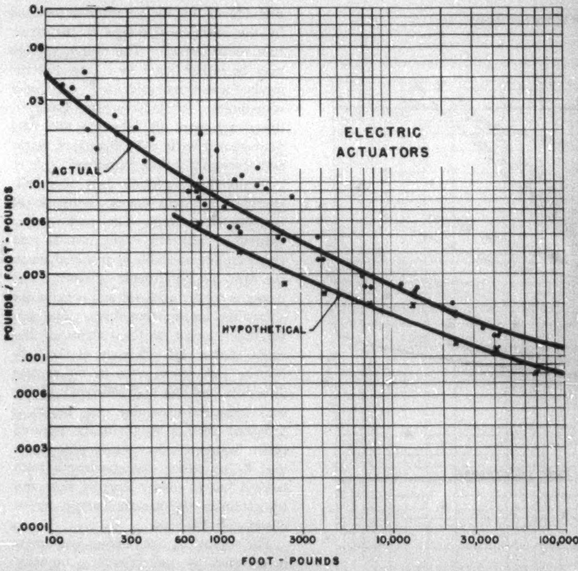


Figure 7 (below). Specific weight of electric actuators

linear electric actuators built by several manufacturers. The smaller sizes utilize an acme thread for the jack screw and the larger sizes, usually over 1,500 foot-pounds, where efficiency is an important factor, use a ball-bearing screw. For any one output rating actuator weights may vary as much as three to one because of a variation in force-stroke combination and variation in operating time. There is, however, a distinct trend and the curve drawn illustrates this clearly. Actuators will vary from this trend if if they have either a very short operating time, necessitating a large motor, or a long operating time, permitting a small motor with a higher gear ratio.

The over-all efficiency of the larger size actuator varies from 40 to 70 per cent. The smaller sizes will be below these values because over-all efficiency is not important when compared to the weight

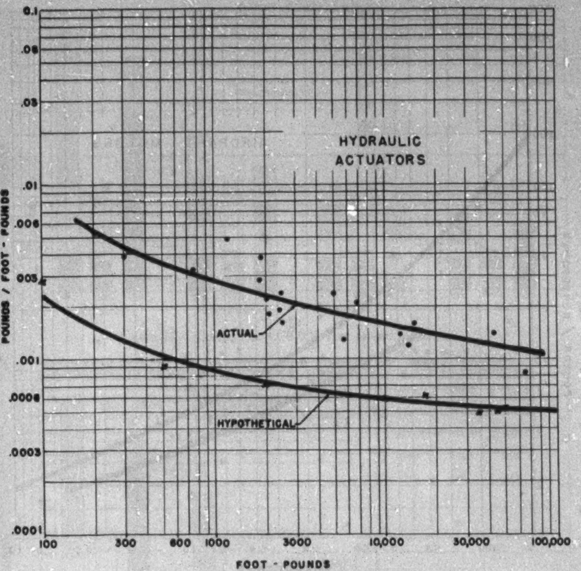


Figure 8. Specific weight of hydraulic actuators

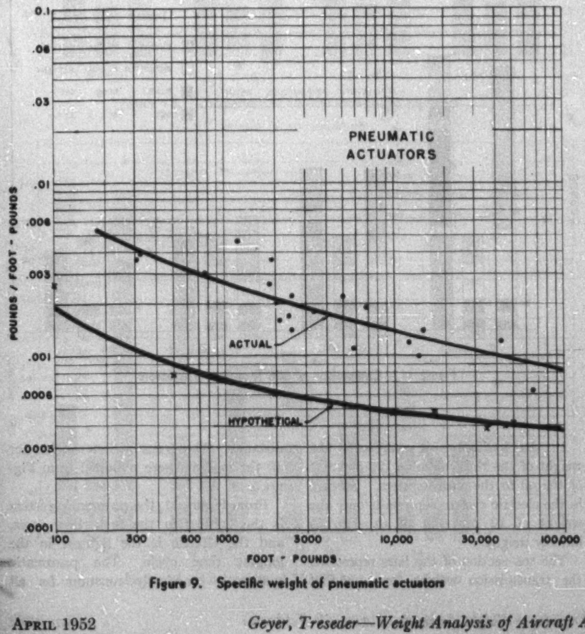


Figure 9. Specific weight of pneumatic actuators

reduction that can be realized by simplified gear boxes, and so forth.

Some weight penalty is experienced when the actuator must fit a complicated structural assembly and when the manufacturer must effect some standardization of motor and screw sizes. To find out how much weight could be eliminated by design alone, a series of eight hypothetical actuators were designed. In each case the motor screws were selected as optimum and the actuator was not used as a structural member. The line through these points represents the best that can be done with available design data.

The specific weight was next compiled and plotted in Figure 8 for a series of 20 hydraulic actuators of various sizes. There was less spread than in the case of the electric actuators because the size of a hydraulic actuator is independent of its rate. The values given are for 1,500-pounds-per-square-inch pressure units designed for a particular installation and with production units. As in the case of electric actuators, eight hypothetical units were designed. No effort was made to facilitate manufacturing. Material was working at allowable stress limits, the pressure was increased to 3,000 pounds per square inch, and the unit was not used as part of the aircraft structure.

Figure 9 shows the same relationship for pneumatic actuators. The pneumatic actuators are simply hydraulic actuators used with compressed air rather than oil and do not have provisions to limit travel rate. Actuators designed specifically for pneumatic use will be somewhat heavier than shown in Figure 9.

Discussion of Bar Charts

The comparison of the three types of systems is shown in the bar charts of Figures 11 to 14 inclusive. Four sizes of actuators were chosen. To show the effect of different rates of operation, each size was computed for four rates. Both sizes and rates were picked hypothetically to cover the entire field of possible operation.

In comparing the data in the following charts, we will not deal with actual figures but just point to the highlights and trends of the over-all picture. Figure 11 is typical of the bar charts. The code block at the top of the chart shows that the bottommost section of the bars represents the actuator weight plus the weight of operating medium. For the electric actuators the specific weight of an actuator for an 18-second operating cycle was taken from Figure 7. The horsepower of this actuator was then computed, assuming an over-all efficiency of 40 per cent for the 600 foot-pound, 50 per cent for the 3,000 foot-pound, 60 per cent for the 15,000 foot-pound, and 70 per cent for the 75,000 foot-pound. Then by using the specific motor weight as shown in Figure 10 for the motor power required at different rates, a correction, either positive or negative, was made to the basic actuator weight. The specific weights of the hydraulic and pneumatic actuators were taken from Figures 8 and 9 respectively. Then, asssuming an actuator efficiency of 95 per cent, the horsepower for each of the operating times was computed.

The next division of bar represents the weight of the generating source. In the electric system, a generator size, arbitrarily chosen four times the demand of the actuator, was picked and the specific generator weight taken from Figure 2. This was multiplied by the sum of the motor power plus line losses to give a value of generator weight chargeable to this actuator. If in the event a four times actuator power demand for the 2-second actuator operation dictated a generator larger than is now being built, the specific weight of the largest size being produced was used. The same specific generator weight was used for each of the same size actuators when operating at a slower rate. It was assumed that 10 per cent of the energy used by the actuator would be provided by the battery so that an allowance was made for this in plotting the generator weight also.

In determining the weight of the generator source in the hydraulic system, the weight of a pump to supply the actuator power, plus line losses, was computed by using the specific weight of a pump four times the demand of the 2-second actuator rate. To this was added the weight of an accumulator capable of supplying 10 per cent of the energy. This same section of the pneumatic bar is just the compressor and is not a direct comparison to the other two systems since the main source is carried in the storage section. The compressor must be just large enough to recharge the storage in a 20-minute period.

The storage weight is represented in the third from bottom section of the bar. This section in pneumatics represents the weight of the air bottles plus the air weight, and was computed on the basis of two complete cycles of actuator operation (per army-navy specification *AN-P-86*). In the hydraulics it represents the weight of the reservoir plus 120 per cent of the oil in the entire system. Storage in the electric system represents only the weight of 10 per cent of capacity in battery weight.

The top section of the bars represents the transmission weight for a 75-foot distance. The transmission weights for all the systems were obtained from Figure 5.

From Figure 11, the pneumatic system is the lightest in the first three cases, and the electric is the lightest in the longest time cycle. The pneumatic system is practically constant for all

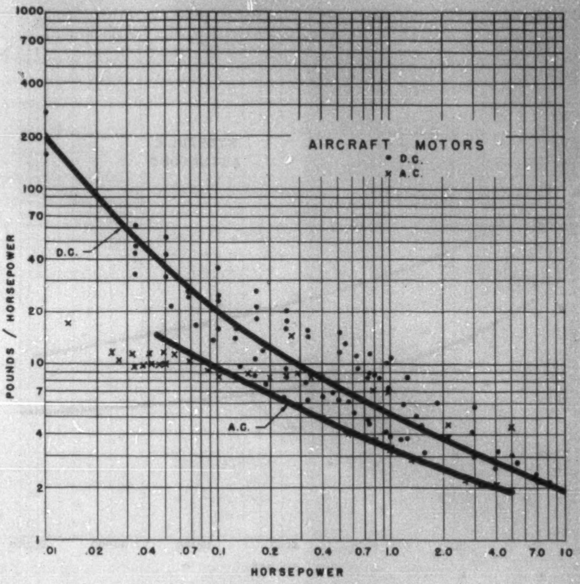


Figure 10. Motor weight

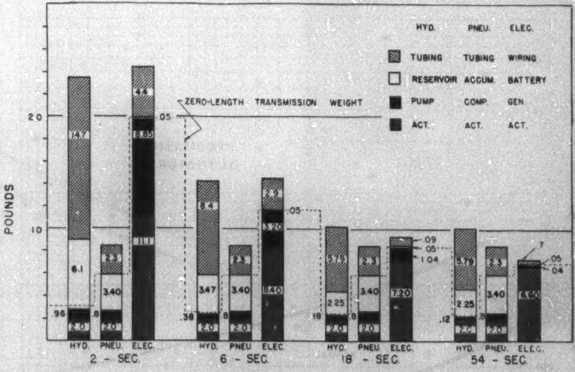


Figure 11. Comparison of 600 foot-pound actuators

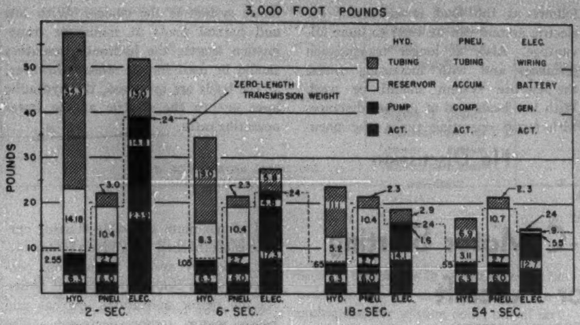


Figure 12. Comparison of 3,000 foot-pound actuators

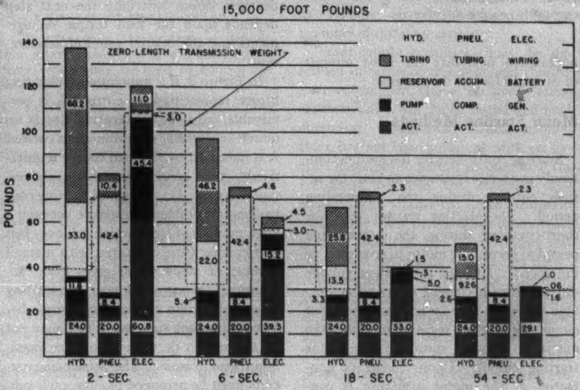


Figure 13. Comparison of 15,000 foot-pound actuators

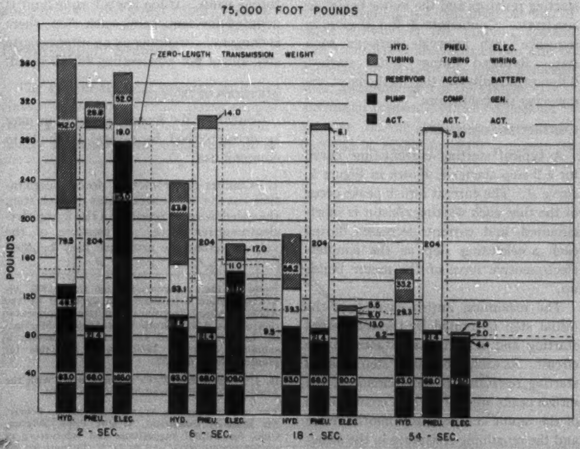


Figure 14. Comparison of 75,000 foot-pound actuators

rates, whereas both the hydraulic and electric start out heavy, and nearly equal, and taper down to be very close to the pneumatic weight at the 54-second case. Hence, in the shortest time cycles the pneumatic system shows a very definite weight advantage, but in the longer times any of the three systems could be used with the slight margin going to the electric. It should be pointed out that the actuator weights for both the hydraulic and pneumatic are the same and constant for all rates. In fact, the pneumatic system is composed of constants, and the bar chart shows it is constant for all rates.

The larger variables in the hydraulic system are in the tubing and storage. In the electric, the larger variables are in the actuator and generator.

The weight distribution comparison will also be true for the bar charts to follow. If present-day practice is continued there would be very little chance for the pneumatic system in the lower rates. And, by the same token, there is very little chance of either the hydraulic or the electric system taking top honors in the fast, remotely located, and lightly loaded actuators.

Since the greatest weight in the hydraulic system and the lowest in the pneumatic and electric systems is in the transmission of power, a line can be drawn through each of the bars in the chart, cutting off the transmission weight and its proportional share of storage weight. This line then shows the system weight at a minimum transmission length. The hydraulic system is then shown to be the lightest system for all rates.

Figure 12, for the 3,000 foot-pound work load at 75-foot operation distance and same rates as considered before, shows that the pneumatic system is the lightest for only the two shorter actuating times. The electric is lightest for the 18 seconds and longer. Another change is that the hydraulic system is lighter than the air system at 54 seconds. A minimum transmission length line has also been drawn on this chart and shows the hydraulic to be the superior system from a weight standpoint for all rates at this distance. Figure 13 shows a 15,000 foot-pound load and the trend is such that the pneumatic is the lightest in the shortest time only. The electric is lightest in the three longer rates. A line for the minimum transmission length shows the hydraulics the lightest for all rates at this load. In the last figure, Figure 14, the load has been increased to 75,000 foot-pounds and we see the electric has gained and now holds three out of four of the times with pneumatic holding only the shortest time. Another surprising thing has taken place in this chart at the minimum line length, and that is the electric system has become the lightest in the longest time.

The trend can be summarized as follows: as the load is increased, the electric system can be used to more advantage. Also, at longer transmission distances and with increasing times, the electric system should be used. With low loads and at greater distances with faster operating times, the pneumatic system is the choice. With low and normal loads at minimum transmission length, the hydraulic actuators should be used for all rates. However, as the loads are increased, the hydraulic loses out to the electric at the lower operating rates.

No Discussion

Reduced Voltage D-C Motor Starting Controllers for Aircraft

H. J. FINISON
MEMBER AIEE

W. C. ANDERSEN
ASSOCIATE MEMBER AIEE

REDUCED voltage controllers are needed for d-c motors for aircraft application to

1. Limit voltage dip during motor starting.
2. Limit inrush current to the motor.

A number of motors currently used on aircraft are large with respect to the power-supply system capacity—in some instances the short time demand of a single motor may equal or exceed the rated capacity of the power-supply system. There have been previous reports of dips in voltage to as low as 25 per cent of normal voltage.[1,2] Although d-c power systems seem capable of recovering from such serious disturbances, large dips in voltage may adversely affect performance of other loads supplied from the system.

Further, where such low voltage may occur, all contactors on the system must be designed with low drop-out voltage. Where only a few motors are the cause of low voltage, it may be better weight economy to equip those few motors with reduced voltage starting controllers and thereby permit application of lighter weight high drop-out voltage contactors throughout.

For lightweight aircraft motors the commutator, brushes, and connections are made as small as possible. Weight savings may also be achieved in these parts if the starting current is limited to reasonable values.

Reduced voltage motor controllers have been used previously in a number of cases on aircraft for particular applications. Frequently, these have been an integral part of a large machine and its controls such as inverters or amplidynes. This paper describes development of a line of controllers that should find wide application flexibility for most aircraft of 24- and 115-volt d-c systems.

Motor Starting Methods

Starting controllers for d-c motors commonly use series resistance to limit the inrush current during the accelerating period. The starting resistor must be short-circuited during the accelerating period to permit the motor to accelerate to full speed. To limit the system voltage dip and inrush current peaks to acceptable values, it may be necessary to short circuit the starting resistors in more than one step. Figure 1 shows a 2-step motor starting circuit. Upon closing of contactor *A*, the motor armature is connected to the line through the series-starting resistors and the motor begins to accelerate. Contactors *B* and *M* close in sequence with intervals between each step. Reduced voltage motor starters differ principally in the methods of sequencing the contactors.

CONTACTOR SEQUENCING

A typical starting current-time curve for a 2-step starter is shown in Figure 2, curve *A*. The current inrush peaks occur at the time each starting resistor is short-circuited, and current decreases during each accelerating step as the counter-electromotive force of the motor builds up.

The maximum inrush current in the initial step is limited by the total series starting and armature resistance in the circuit. As each subsequent step of series starting resistance is short circuited, the inrush peaks are limited by a combination of the motor counter-electromotive force and the remaining resistance in the armature circuit.

The maximum allowable starting current is established for a given motor rating either by system voltage dip considerations or peak motor current limits. As the motor accelerates, current and hence motor torque decreases. The amount which current may be allowed to decrease before switching the next step depends upon the load torque requirements and the desired speed of acceleration.

In Figure 2, if a minimum accelerating torque represented by current I_1 is permissible, only one accelerating step is required (curve *B*). If a minimum current I_2 is indicated by the load torque requirements, two accelerating steps are required (curve *A*). Design of starting controllers is a problem of balance between a large number of steps which achieves fast accelerating time compared with a small number of steps which gives lower accelerating torque and longer accelerating time. Minimum accelerating torque before closing each subsequent step may not be allowed to decrease below the maximum load torque. The sequencing circuits must be arranged to close each subsequent step at the selected limiting torque value. If too low a torque limit is adopted for sequencing each step, there is danger of remaining on an intermediate step and of not accelerating to full speed.

METHODS OF SEQUENCING

Control of d-c motor acceleration may be accomplished by several methods in common use:

1. Counter-electromotive force acceleration—as the motor accelerates on each step, the speed increase results in a rise in counter-electromotive force. A measure of the voltage across the armature may be used

---

Paper 52-134, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE South West District Meeting, St. Louis, Mo., April 15–17, 1952. Manuscript submitted July 12, 1950; made available for printing March 6, 1952.

H. J. FINISON and W. C. ANDERSEN are with the Armour Research Foundation, Chicago, Ill.

The authors express appreciation to G. Scheer, Air Materiel Command, Wright Field, Dayton, Ohio, for helpful suggestions and conducting of special system tests, and to D. E. Bennett for his work on the preliminary system analysis.

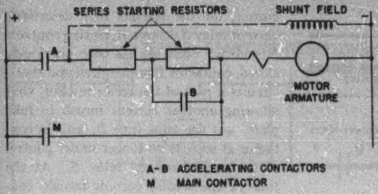
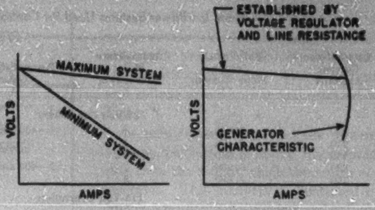


Figure 1 (left). Typical resistance motor starting circuit

Figure 3 (right). General form of power system volt-ampere characteristics



(a) TRANSIENT

(b) STEADY STATE

to determine the condition for closing the next accelerating contactor.

2. Current limit acceleration—as the motor accelerates on each step, the current decreases from a peak to a lower value as the counter-electromotive force of the motor builds up. The current may be used as a criterion for determining the condition for closing the next accelerating contactor.

3. Definite time acceleration—a definite time interval obtained by the use of time-delay relays may be used to control the accelerating contactors.

4. Combination of these methods—either the counter-electromotive force or current limit method may be combined with definite time acceleration.

Definite time acceleration, by itself, is not suitable for aircraft application because of the wide differences in inertia of the loads to which a given motor may be applied and also wide differences in power-supply conditions.

The counter-electromotive force and current limit methods of acceleration both provide quantities which may be sensed and used as criteria to control the accelerating contactors. Both methods have been investigated to determine their suitability for application to motors operating from aircraft power supply systems.

Power-Supply Systems

The system engineer working with aircraft systems must consider power-supply conditions which vary greatly. On a particular airplane power supply may differ because of changes in engine speed or because of one or more generators on a multigenerator system being out of service. In addition, particular equipment, such as motor controllers, may be applied to various aircraft having differing amounts of installed capacity.

Power-supply conditions may be described in terms of transient and steady-state volt-ampere characteristics.[3] These characteristics will define initial and final operating conditions. Unfortunately, there has not been sufficient work on aircraft machines to permit prediction of the transition between the transient and steady-state conditions. Also, transient and steady-state performance is not available for many of the power-supply machines and systems now commonly used.

Characteristics of both the maximum and minimum power supply conditions expected to be encountered are required for the design of each controller rating. The transient and steady-state volt-ampere characteristics of the maximum system were specified as giving a 1/2-volt drop per unit of rated current on the 24-volt system and a 2-volt drop per unit of rated current on the 115-volt system. The transient characteristics can also be expressed in terms of the initial available short-circuit current of the system. These maximum transient values correspond to systems having available short-circuit currents of 12,000 to 19,000 amperes for the 24-volt system and 3,000 to 5,000 amperes for the 115-volt system. Volt-ampere characteristics of the minimum system were determined from performance data on the power-supply sources shown in Table I.

The power-supply system is assumed to have adequate steady-state capacity and the voltage regulators to be sufficiently rapid in response such that voltage does not dip below the initial transient value. Figure 3 illustrates the general form of transient and steady-state characteristics for the supply systems as used for the design analyses.

Counter-Electromotive Force Acceleration

The basic circuit for counter-electromotive force acceleration control is shown in Figure 4. When the control switch is closed, contactor *A* is energized and closes contact *A*. Contactors *B* and *M* sense armature voltage as the voltage increases. Voltage setting resistors R_V are adjusted to allow contactors *B* and *M* to operate at respectively increasing values of armature voltage. In the final step, contactor *M* operates to short-circuit the balance of the starting resistance, thus connecting the motor across the line and de-energizing the accelerating contactors. An auxiliary relay may be advisable in place of the auxiliary contacts on contactor *M*.

The chief limitation of the counter-electromotive force starting circuit involves the question of how satisfactory the armature voltage serves as a measure of motor torque. The armature voltage E_a, which is used for control, is deter-

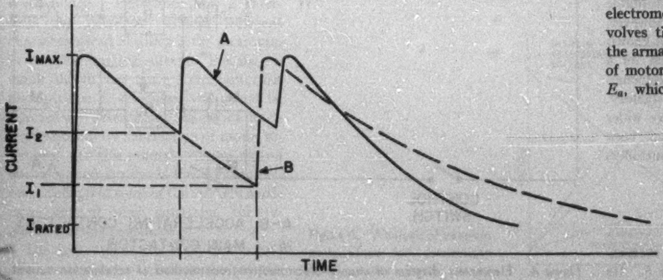


Figure 2. Typical current-time curves during motor starting

Curve A—two step starter
Curve B—one step starter
Both curves for the same motor and power supply system

Table I. Power Systems Used for Controller Designs

Rated Motor Current Amperes	Full Voltage Inrush Motor Current	Approximate Motor Horsepower	Power Supply Sources for Minimum Systems	
			Generators	Battery
24-Volt Controllers				
200	2,000	4	1—400 amperes	1—34 ampere-hours
240	2,400	5		
280	2,800	6		
320	3,200	7	2—400 amperes	1—34 ampere-hours
360	3,600	8		
115-Volt Controllers				
50	500	4		
60	600	5		
70	700	6	2—29-kw 167-ampere generators	None
80	800	7		
90	900	8		

mined by system voltage as well as armature current

$$E_a = E_L - I_a R_s$$

where

E_a = armature voltage
E_L = supply system voltage
I_a = armature current
R_s = series current limiting resistance

If motor torque is a function of armature current I_a, armature voltage E_a, may not measure motor torque satisfactorily if system voltage E_L differs between operations.

Since aircraft operate with varying and limited power supply capacity, we cannot assume that E_L will always recover to full system voltage during the accelerating period. Even with large power-supply capacity we cannot assume that a system will recover full voltage before motor current and torque have dropped below a specified minimum value. This may be particularly true for series motors driving low inertia loads.

There is some question whether dropping below a specified minimum torque during a rapid acceleration is a serious problem. However, it is always possible that in the interest of economy power-supply capacity may be limited and system voltage may not recover to full value during motor acceleration. Hence armature voltage may not be an acceptable measure of minimum torque. However, it is probable that with sufficient knowledge of the power-supply system for a particular application, counter-electromotive force control would be satisfactory. Armature current appears to give wider application flexibility, and has been adopted for the line of starting controllers being described.

Current-Limit Acceleration

The circuit diagram of a 2-step current-limit starter is shown in Figure 5. Its operation is based on the actuation of current control relay S, which is sensitive to a voltage directly proportional to the motor starting current. This relay controls the sequence of operation of the stepping and main contactors, B and M. When the control switch is closed, line voltage is applied only to contactor A which closes to energize the motor circuit through the series resistors. The closing of auxiliary contact A_1 energizes contactor B through relay contacts S_1. However, before contactor B can close, it is de-energized by opening of contact S_1 as the current relay picks up due to the motor inrush current.

As the motor accelerates the starting current decreases from its inrush value. When motor current decreases to a selected minimum (200 per cent of rated current for the starters here described) current relay S drops out, closing contacts S_1 and S_2. Since A_1 has been previously closed, contactor B is energized and short-circuits a part of the series resistor, thus allowing another current inrush to take place. At the same time B_1 seals in contactor B so it is no longer under control of the current-control relay S. At the time of the second current inrush, contactor M is energized momentarily but does not pick up because the inrush current again energizes relay S and opens S_1 and S_2. When the current decreases during this second accelerating step, relay S again drops out at the preselected current, thus closing S_1 and S_2 allowing contactor M to pick up and connect the armature of the motor directly across the line. The starting controller is now in its continuous operating condition and main contactor M is sealed in by auxiliary contact M_1. Accelerating contactors M and B are de-energized during continuous operation by the opening of auxiliary contact M_2.

Design Analysis

The circuit operation and design requirements may be analyzed from a study of the current voltage relationships as shown in Figure 6. The steady-state and transient system characteristics can be represented as previously shown in Figure 3. During the current inrush period, the motor will operate on the system transient characteristic, and during the accelerat-

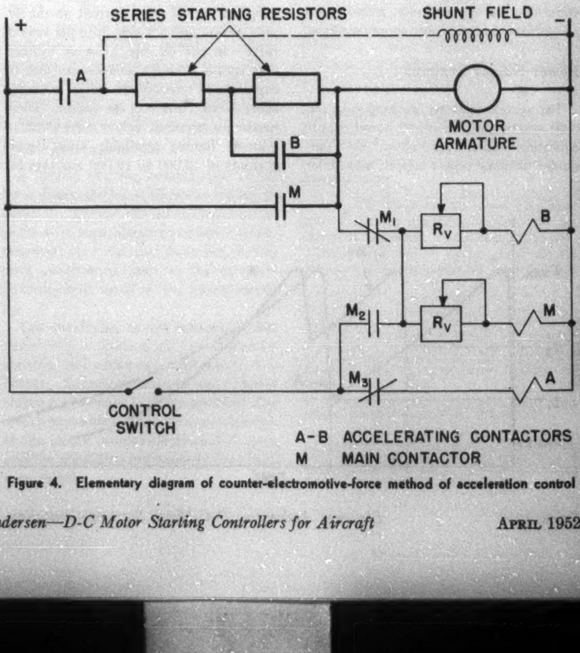


Figure 4. Elementary diagram of counter-electromotive-force method of acceleration control

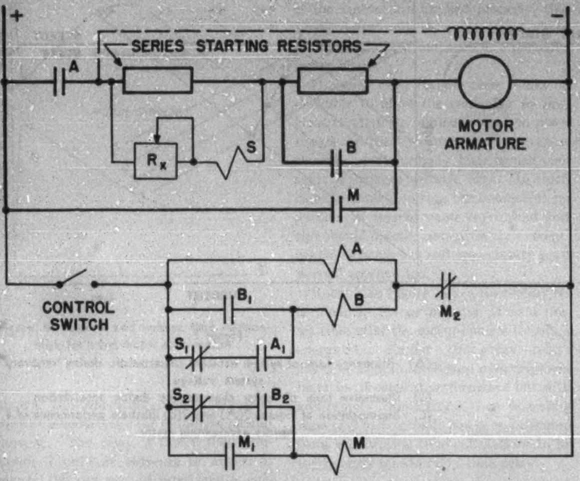


Figure 5. Elementary diagram of current-limit method of acceleration control

ing period will usually recover to and operate on the steady-state system characteristic.

Figure 6 illustrates the method of analysis for transient conditions for both minimum and maximum systems.[3] The intersection of the motor armature resistance line with the system transient characteristic defines the across the line starting current and voltage. For example, points *C* and *D* define the initial operating conditions on the minimum and maximum power-supply systems. An excessively high current would be obtained on the maximum power-supply system and an excessively low voltage would be obtained on the minimum power-supply system.

Resistance line *1* intersects the minimum system transient characteristics at *A* which is the minimum permissible system voltage. Resistance line *2* intersects the maximum system transient characteristic at *B* which is the maximum permissible starting current. For the example shown, resistance *1* represents the higher value of resistance and would be selected for determining the series starting resistance by subtraction of motor resistance. In this example voltage dip on the minimum system is the limiting factor, although in some of the starters maximum current may determine the required starting resistance.

The next step in the analysis involves an investigation of performance as the system voltage recovers and the motor accelerates. Exact methods of analysis have not been developed yet but a qualitative evaluation will aid in understanding the problem.

It is probable that the recovery of the system voltage may be represented graphically by curves parallel to the initial transient characteristic as shown in Figure 7(A). Similarly, as the motor accelerates, the counter-electromotive force rises and the motor characteristic can be shown by a series of curves parallel to the starting resistance characteristic, Figure 7(B). By superposition of these two figures, the current-voltage pattern which will be followed with time is obtained as in Figure 7(C). After voltage recovers to steady-state value, further reduction in current will depend on the magnitude of load torque. In the example shown in Figure 7 the rate of motor acceleration is of the same order of magnitude as rate of recovery of system voltage and current-time plots of current and voltage would appear as illustrated in Figure 8.

In many cases, however, the system voltage recovers at a much more rapid rate than the rate at which the motor accelerates. This analysis is illustrated in Figure 9 resulting in a current and voltage pattern which would appear as shown in Figure 10. Here the motor current rises above the initial transient value and this maximum current may be the consideration in defining the required series starting resistance.

Figure 11 shows the complete analysis for determining resistance values for steps that follow the first. Point *1* illustrates the initial operating condition with two steps in series with the motor armature. As counter-electromotive force increases during motor acceleration and as the system voltage recovers to steady-state conditions, the current value progresses to point *2* and toward point *3* along the steady-state characteristic. At point *3*, which is the minimum acceptable motor current, the counter-electromotive force has built up a value determined by the intersection of the total resistance line on the voltage scale at point *4*. At this time a portion of the total resistance is short-circuited by an accelerating contactor to allow another current inrush. The amount of resistance left in the circuit is determined by a resistance line drawn from point *4* at the maximum current allowed on the transient system characteristic. In many cases where the rate of recovery of system voltage is much more rapid than is the rate of motor acceleration, the intersection of the resistance line with the steady-state characteristic should be used for design where current is the limiting design consideration.

For the case shown in Figure 11 using the transient characteristic the second resistance step may be determined by re-

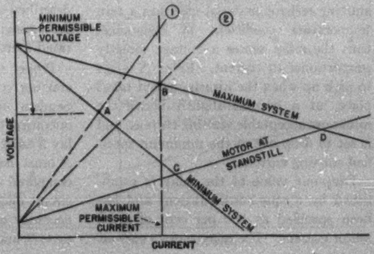


Figure 6. Analysis of transient performance for determining initial motor inrush current

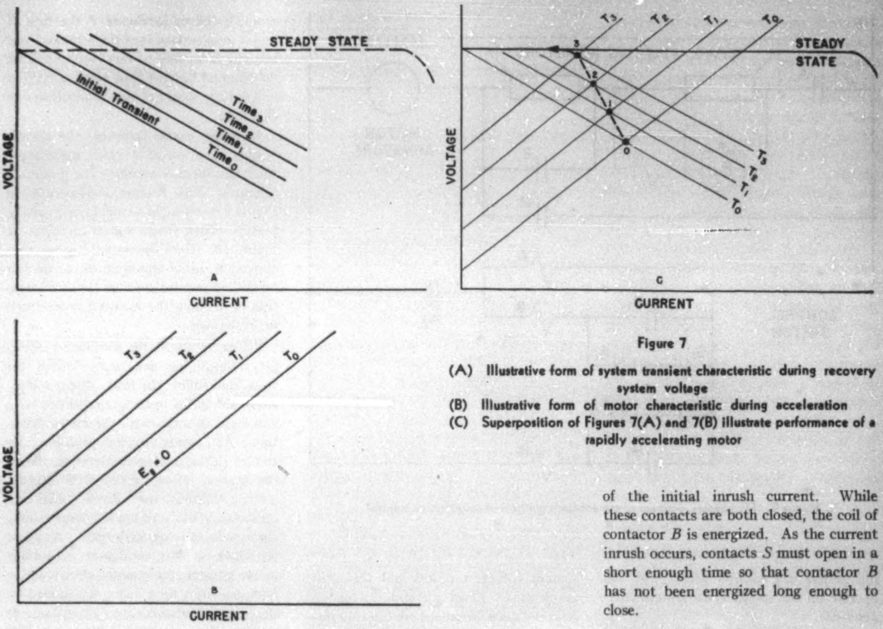


Figure 7

(A) Illustrative form of system transient characteristic during recovery system voltage
(B) Illustrative form of motor characteristic during acceleration
(C) Superposition of Figures 7(A) and 7(B) illustrate performance of a rapidly accelerating motor

sistance line R_a+R_{s2} through points *1* and *3*. The inrush current for this step will rapidly follow some path between points *3* and *1*. The fact that the resistance line (R_a+R_{s2}) has a greater slope than the motor armature resistance line indicated another resistance step is required before the motor is connected across the line. If the resistance line R_a+R_{s2} had had a slope equal to or less than the slope of the motor armature R_a, only one accelerating step would have been necessary.

Operation of Current-Sensing Circuit

Actuation of the accelerating contactors is controlled by relay *S* Figure 5, connected across the starting resistor. The starting resistor material used has a zero temperature coefficient of resistance; thus the relay senses a voltage directly proportional to current. Relay *S* is set to pick up when the inrush current takes place, at a current slightly above the minimum acceptable starting current, and is set to drop out at the minimum specified starting current.

Drop-out value of the relay is established by torque considerations and has been specified as 200 per cent of rated motor current. The relay pickup current will, of course, exceed drop-out value. For example, it might be 300 per cent of rated motor current as compared with dropout of 200 per cent of rated motor current. With a relay set to pick up at 300-per-cent current, the relay will not be actuated at lower values of current and the second stepping contactor will close without awaiting acceleration of the motor. Some acceleration, of course, will occur during the inherent closing time of the stepping contactor. However, such low values of current will occur only on a system of limited capacity and no difficulties will be evident, provided that on closing of the second step voltage does not dip below the specified minimum. Analysis of the systems for the controllers being described demonstrated that pickup currents ranging from 310 to 380 per cent of rated motor current would be acceptable.

Proper operation of the current limit starting control circuit is primarily dependent upon the relation between the operating times of the current-control relay *S* and the contactors *B* and *M*. Relay *S* must have a short operating time as compared with contactors *B* and *M*. The circuit diagram of Figure 5 indicates the necessity of this requirement, since contacts S_1 and A_1 are closed at the beginning of the initial inrush current. While these contacts are both closed, the coil of contactor *B* is energized. As the current inrush occurs, contacts *S* must open in a short enough time so that contactor *B* has not been energized long enough to close.

Co-ordination of Relays and Contactors

Typical curves of contactor closing time as a function of applied voltage are shown in Figure 12. These curves indicate a minimum closing time of the order of 20 to 30 milliseconds at rated voltage, with operating time increasing with a decrease in applied voltage. (Temperature effects on contactor closing time were found to be negligible.) Minimum contactor closing time is compared with the operating time of the current control relay.

Operating time curves for three current control relays are shown in Figure 13. Each relay has been preset to drop out at 200 per cent of rated motor current, and to pick up at 225 per cent of rated motor current. (Drop out of these relays is about 90 per cent of pickup value.) The operating time has been plotted as a function of rated motor current. The vertical constant time lines are the closing time for three different contactors at rated voltage.

To obtain satisfactory circuit operation the current control relay must pick up in a shorter time than does the contactor (at 310 per cent of rated current). Relay *a* was reported to have an operating time of less than 5 milliseconds; however, as

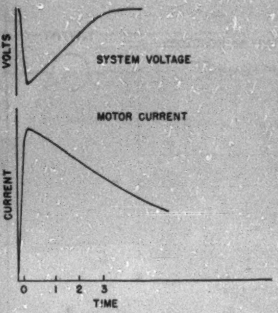


Figure 8. Current and voltage time relationships for a fast motor acceleration

shown, this operating time is obtained only at several times the minimum pickup current. The relay is faster than contactor *1* only at currents in excess of nearly 400 per cent of rated motor current. Relay *b* has a closing time shorter than contactor *1* at all values of current above 240 per cent of rated motor current. This gives an approximately 2 to 1 margin at 310 per cent of rated motor current and hence this relay would probably be suitable for the current control service. However, relay *c* has a very fast operation—less than 2 milliseconds—even at its minimum pickup, and hence substantially avoids all problems of time coordination.

CURRENT-TIME CONTROL

In some cases load torque may not drop sufficiently low for current to drop below 200 per cent of rated motor current. If the motors are applied properly, this would occur only in cases where there is abnormal operation, such as a partially frozen load.

However, even in these cases it may be desirable to allow the controller to progress through its sequence steps to place the motor directly across the line, even at the risk of abnormally high inrush current or system voltage dip. In such cases, a control to step the contactors regardless of current value is required, but this should operate only after the current-control scheme has had opportunity for a normal acceleration.

Relay *T* in Figure 14 is a time-delay relay, which closes contactor *M* at a preset time after the motor-control circuit is energized. Similar time-delay relays could be used on each step with improvement in theoretical performance but with additional complexity. An operating time of 2 to 5 seconds has been indicated from accelerating time calculations to be satisfactory for the relay-time delay.

Current-Limiting Resistors and Accelerating Time

Both the current during acceleration and the total accelerating time influence the design of the series starting resistors. The accelerating time of a d-c motor is dependent upon the value of driving torque acting upon the combined inertia of the motor and its load.

INERTIA CONSTANTS

Calculation of accelerating time requires a knowledge of the motor and load inertia constant. Representative values of wk^2 for open continuous d-c aircraft motors are shown by Figure 15 for 7,500-rpm motors. For other speeds, the wk^2 value will be different since the armature size will be different. However, there is a tendency for the total accelerating time to remain approximately constant for motors of differing speed ratings since accelerating time is proportional to the product of wk^2 and the final operating speed

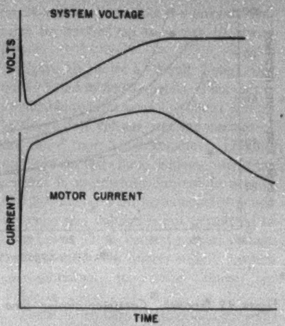


Figure 10. Current and voltage time relationships for a slow motor acceleration

$$t = \frac{wk^2\,\text{rpm}}{308T}$$

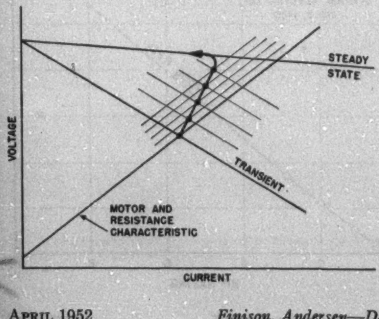
where

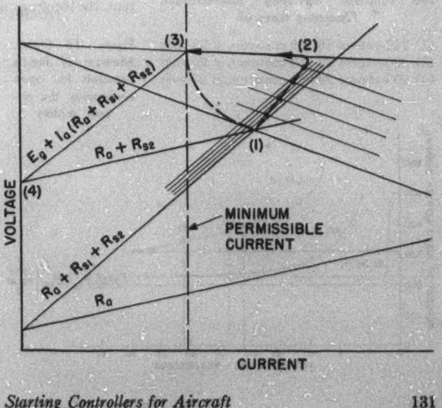
t = accelerating time, seconds
wk^2 = inertia constant, pound-feet^2
rpm = motor speed
T = accelerating torque, pound-feet

For accelerating time and resistor design calculations, a value of three times the motor wk^2 was assumed to be the maximum combined inertia of the motor and load for any application. In general,

Figure 9 (below). System performance during starting of a motor having a slow rate of acceleration

Figure 11 (right). Design analysis for determining starting resistance values





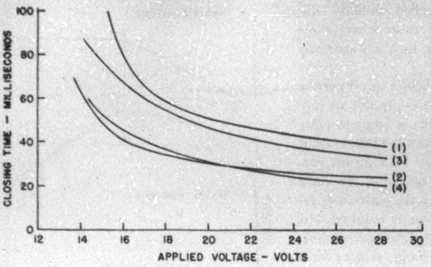


Figure 12 (above). Contactor closing time

(1) 600-ampere 28-volt contactor
(2) 300-ampere 28-volt contactor
(3) 400-ampere 28-volt contactor
(4) 200-ampere 28-volt contactor

Tests made at approximately 24 degrees centigrade

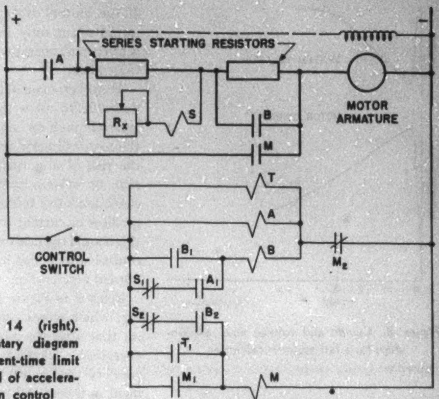
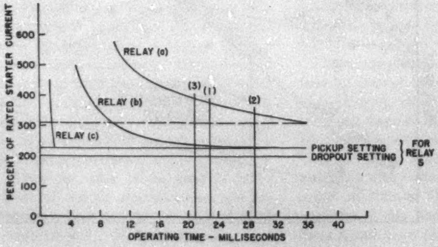


Figure 14 (right). Elementary diagram of current-time limit method of acceleration control

the inertia of the load probably will not exceed the inertia of the motor itself. For example, in a motor-generator installation, the generator has the same speed and approximately the same physical size as the motor; hence its inertia would be equal to that of the motor. In this case, the combined inertia of the motor generator would be twice the inertia of the motor itself. For a motor driving a load such as the landing gear or wing flap, the large speed reduction between the motor and the load in most cases gives a relatively low load inertia when referred to the motor speed, although additional study of such loads would be desirable.

Figure 13 (below). Co-ordination of relay and contactor operating characteristics. Operating times of:

(1) 300-ampere 28-volt contactor, at 28 volts
(2) 300-ampere 28-volt contactor, at 20 volts
(3) 200-ampere 28-volt contactor, at 28 volts



On the basis of the preceding considerations, a total of three times the representative motor wk^2 was chosen for calculating accelerating time.

Motor Torque

The torque in a d-c machine is proportional to the field excitation and the armature current

$$T = K\phi I_a$$

where

T = torque
ϕ = field flux
I_a = armature current
K = constant

For a shunt motor operating with constant field current, the torque available is proportional to armature current

$$T = K_1 I_a$$

For a series motor, where field excitation is proportional to the armature current, the torque available is

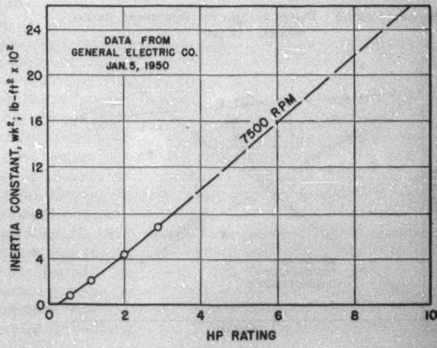
$$T = K_2 I_a^2$$

However, aircraft machines have small field structures which may saturate at relatively low current values and starting torque may not be greatly in excess of that indicated by direct proportionality as shown for the shunt motor.

For shunt motors, where the field is energized at the same time that the initial contactor closes, full field flux and hence full torque are not immediately available due to the inductance of the field circuit. Tests on a typical 9- and 11-horsepower shunt motor showed field time constants of approximately 0.07 second. Because of saturation, the time required to reach 63 per cent of normal flux would be somewhat less than 0.07 second.

The influence of the field build-up should be considered in evaluating the available torque to assure that satisfactory performance will be obtained with shunt motors. In some cases, with limited power-supply systems, it is advisable to

Figure 15 (right). Moment of inertia constants for open continuous d-c aircraft motors



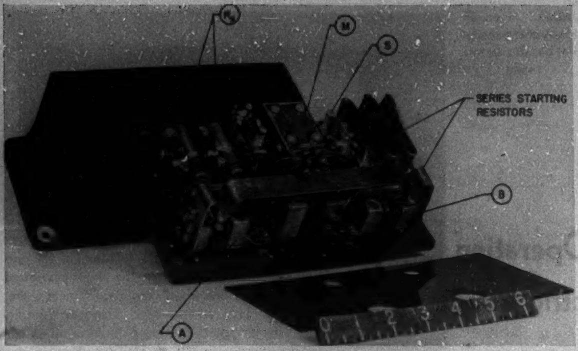


Figure 16. Twenty-four volt reduced voltage starting controller rated at 360 amperes, approximate weight 8 pounds

energize the shunt field some time prior to the time the armature is energized in order to give opportunity to build up field flux with full system voltage available. This aids in obtaining high breakaway torque.

With reduced voltage starters on shunt machines, it is important that the shunt field be connected on the line side of the starting resistor. If it is connected on the motor side of the resistor, only part of the system voltage will be available for the field and adequate starting torque will not be obtained.

For design purposes, motor developed torque was assumed proportional to armature current, and a friction torque equal to rated torque was deducted to obtain the available accelerating torque.

Calculations of accelerating performance were made by taking small increments of current change, and obtaining corresponding change in speed, average torque, and incremental accelerating time. From these calculations the total watt-seconds of heat developed in the starting resistance were obtained, and resistor designs completed. Due to the short accelerating time—in the order of 0.5 second—it was assumed that all the heat produced in the starting resistor contributed to elevating the resistor temperature. That is, no heat was dissipated during the accelerating period.

Motor Controllers

The line of motor starting controllers developed includes ten different ratings, as summarized in Table I, for application on 24-volt and 115-volt d-c aircraft systems for motors up to about 10 horsepower.

Figure 16 shows a 24-volt 2-step controller rated at 360 amperes. This unit is applicable to either series or shunt motors having a full voltage starting current of not more than 3,600 amperes. Laboratory system tests have demonstrated that current is successfully limited to a peak of 1,200 amperes on high-capacity power systems, and voltage is not permitted to drop below 60 per cent of normal voltage on the minimum power-supply system corresponding to those shown in Table I.

These controllers are intended for nonreversing service. Where reversing drives are encountered, two of these units may be used. It is probable that some saving in total weight may be achieved by developing a similar line of controllers for reversing service.

In applying the controllers, circuit protection and some form of overload protection will be required external to the starting controller. Future development may indicate the desirability of incorporating short-circuit protection and even possibly motor overload protection into a combination motor controller.

References

1. DESIGN PRINCIPLES FOR ELECTRICAL LANDING GEAR CONTROL SYSTEMS, H. J. Finison. *Aeronautical Engineering Review* (New York, N. Y.), June 1945, pages 5–19.

2. ELECTRICAL SYSTEM TESTS—B-29 AIRPLANE, H. J. Finison, R. H. Kaufmann. *Technical Report Number 5812*, United States Army Air Force (Washington, D. C.), 1945, page 11.

3. D-C POWER SYSTEMS FOR AIRCRAFT, PART V, R. H. Kaufmann, H. J. Finison. *General Electric Review* (Schenectady, N. Y.), May 1946, pages 46–54.

No Discussion

Maintenance and Operating Experiences—Diesel-Electric Locomotives

C. H. GRIFFITH
APPLICATION PENDING

E. H. PFEIL
NONMEMBER AIEE

Synopsis: This paper outlines certain difficulties experienced with diesel-electric locomotives on the Missouri-Kansas-Texas Railroad since the first road service units were placed in operation during 1947. With four different makes of locomotives, totalling 197 units operating on 3,242 miles of railroad, furnishing the basis for the observations outlined, it can be assumed that similar difficulties have been prevalent on other railroads of comparable or greater size and diesel ownership. Any criticism offered should be considered constructive as it is given sincerely and with a hope that it may be of some benefit to manufacturers and users alike.

Main Generator—Commutation

WITH one group of 21 road freight units starting service in June and July 1947, there was a marked pattern of commutator bar discoloration followed by bar etching and flat spots which necessitated the resurfacing of the commutators on all units. This trouble developed when locomotives had been in service 5 to 6 months and had operated approximately 80,000 miles. It is believed that it was caused by the improper functioning of automatic transition and possibly aggravated by commutator eccentricity.

Since the commutators were resurfaced and automatic transition modified so as to eliminate the cycling between series parallel shunt and parallel motor connections, no similar trouble has been experienced, the commutators now being in excellent condition after more than 700,000 miles of operation. Another modification made on these units was a change from closed circuit to open circuit transition. At the same time current values for controlling backward transition were increased approximately 500 amperes, thereby reducing the number of motor connection transfers; these changes, it is thought, were an aid in the elimination of commutator trouble.

On another group of 18 road freight units placed in service from November 1948 to March 1949, there was considerable trouble in maintaining commutator surface due to threading and copper drag which tended to produce flashovers. These generators had commutators of the open riser type which were difficult to keep clean principally because the air drawn through the generator by the generator blower was laden with oily vapor collected in the air stream in passing over the top of the engine. These oily vapors made it difficult to keep the front and rear "V" ring string bands clean, thereby producing a condition which was contributory to flashovers.

These conditions have been improved by changing the grade of brush, reducing the number of brushes per holder to obtain greater current density, and blocking some car body filters to change the air flow pattern in the engine room. In this case it is felt the first two changes mentioned contributed most to relieving the trouble.

On a third group of 20 road freight units placed in service during July 1949, there have been quite a few flashovers which are attributed principally to oily vapors being blown through the main generator and collecting on brush holder insulators. These oily surfaces collect carbon dust and thereby furnish a path to ground. These particular units are of the same manufacture as the first group mentioned, but are of a later design and model. In the modernization of these units the manufacturer reduced the number of car body filters $33^{1}/_{3}$ per cent and applied louvres which gave the locomotive a much better external appearance. The filters eliminated were at a location in the end of the car body where the major portion of the engine room air requirements are, namely, engine air, generator blower, front traction motor blowers, and air compressor. With the reduction in the number of filters, the restricted air flow through the louvres, and the revision of filter location, it is evident that some of the generator blower air must be supplied from the air stream passing over the top deck of the engine.

Recently there was a fire on one of these units in the main generator, a fire which might be classified a trash fire as inspection disclosed the fact that the burning material was an accumulation of grease which evidently had been ignited by a severe flashover and fanned by the generator blower. Fortunately, the damage was negligible, as a maintainer who was riding the locomotive took the unit off the line, shut the engine down, and extinguished the blaze with fire extinguishers. When the generator was removed and thoroughly cleaned with ground corn cobs no serious damage was found, merely the burning of approximately 1 square inch of outer insulation on a compensating field connection.

These oil laden vapors produce a gummy condition in the brush holders which in turn is apt to set up ring fire and eventually a flashover between brush holders of opposite polarity.

To alleviate these troubles a program of installing six additional standard car body filters in each unit located in the area of the maximum air demand has been initiated.

With this particular type of locomotive the main generators are being operated at greater temperatures than those of the older type as a considerable quantity of

---

Paper 52-145, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE South West District Meeting, St. Louis, Mo., April 15–17, 1952. Manuscript submitted January 9, 1952; made available for printing February 27, 1952.

C. H. Griffith and E. H. Pfeil are both with Missouri-Kansas-Texas Lines, Parsons, Kans.

air is being recirculated through the generator.

Figure 1. Cabinet with two twin graphic instruments and telephone

Switch Gear and Control Equipment

There have been many locomotive unit failures on one class of locomotive due to poorly designed contacts on the switch used for the transfer of motor connections from power to braking. These contacts cannot be made trouble-free with any reasonable amount of shop maintenance and have been responsible for many train delays and tonnage reductions.

Another item that has caused much trouble from a maintenance standpoint is the design of interlock contacts. On one make of locomotive, trouble from interlock contacts has been negligible while on a competitive make of locomotive this very important item has been the cause of numerous delays and costly maintenance.

A third problem that has caused concern on the part of maintenance forces is the type of equipment used for transition control. On one make of locomotive, control is by use of an axle generator which transmits a wheel speed signal to three or more relay panels which control motor transfer switches. A special drive and test meter are required for satisfac-

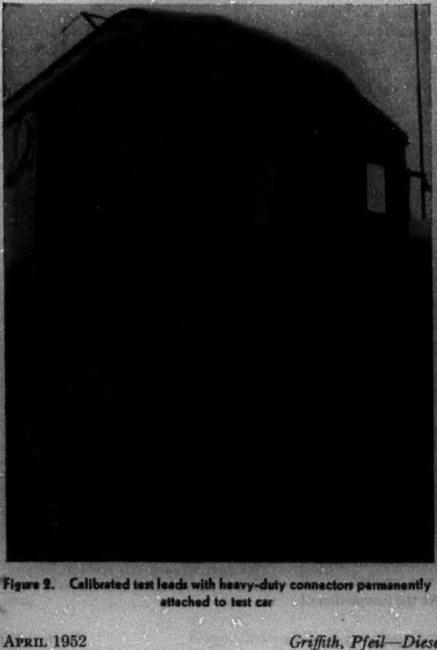


Figure 2. Calibrated test leads with heavy-duty connectors permanently attached to test car

Figure 3. Test cables connected between test car and locomotive

tory checking of this equipment. With all of this, maintenance often is reduced to guesswork since road operation is difficult to duplicate in the shop.

As a contrast, equipment performing similar control functions on another make of locomotive has been reduced to three rather simple relays. Special equipment for testing in this case involves a small motor-generator set and voltmeter. With this equipment, very few cases of trouble have been experienced that cannot be corrected definitely in the shop.

The point is not to champion any one make of equipment, but to bear in mind that from a railroad maintenance point of view the simplest method of accomplishing a given result is generally the best. It is felt that this is true even though the final accuracy of the simpler method may not be as close as a more involved and costly method. Much help also could be given maintenance forces by the manufacturers in prompt correction of recognized faulty designs.

Control equipment on some locomotives is installed in such a compacted mass that it is almost impossible to reach some devices or terminals to clear faults which develop. These places are not necessarily inaccessible during the construction at the manufacturing plant, but are rendered so when the hood or housing is placed over them. If design engineers had the responsibility for maintenance, it is felt they would give more thought to the accessibility feature.

Traction Motors

On one particular make of traction motors it is necessary to grind commutators at 150,000 to 200,000-mile intervals while on two competitive makes a similar condition does not exist. These commutators when checked with a dial indicator show variations up to 0.012 inch and vary between the three brush tracks. Commutator surfaces show a series of flat spots which may span several bars and very seldom extend across the entire commutator face. The manufacturer attributes this condition to stall burns, but we cannot understand why similar conditions are not prevalent on other makes of locomotives operated by the same personnel in identical service unless it is because of differences in load control methods.

On-the-Road Testing with Graphic Instruments in a Test Car Coupled to the Locomotive

The car just mentioned is an all steel passenger car fitted for this particular use and has an instrument or office room, sleeping accommodations for nine people, kitchen, dining room, and a room for the attendant.

The forward end of the car contains a cabinet, with desk attached, in which are mounted two twin graphic instruments, generally operated at a chart speed of 12 inches per hour, containing the following elements (see Figure 1):

A wattmeter for recording main generator output.

A speed recorder operating from generator, belt driven from car axle.

A voltmeter for recording main generator voltage.

An ammeter for recording traction motor current.

Each instrument is fitted with two chronograph pens, one on each being operated simultaneously by a pushbutton and generally used for marking mile posts on the margin of the charts. Two of the other chronograph pens record service and emergency brake applications.

A telephone headset and chest transmitter is located at the desk in front of the instruments and is used for communication with one or both cabs of the locomotive by the use of portable telephones.

Test leads, calibrated with the instruments, are in two parts and fitted with heavy-duty connectors. The portions attached to the instruments are permanently installed in the car and lead out through the vestibule for convenient connection to the portion of the cables used on the locomotive. All road locomotives are fitted with openings in the nose section for the accommodation of the locomotive portion of the cables which are secured with strain relief bushings and special lock nuts; see Figures 2 and 3.

With these instruments an undesirable feature in generator control was found which caused the generator to operate at subnormal rating for $7^{1}/_{2}$-minute periods during the process of making forward transition on traction motors. When this was brought to the attention of the manufacturer's representatives, they said they had delivered large quantities of similar locomotives to larger roads and never had received similar complaints from them. This trouble was overcome by their engineers, but would probably have been passed unnoticed if a complaint had not been registered.

In another case there was trouble with cycling from transition 2 to 3 and back to 1. To correct this trouble the manufacturer's engineering department made a modification which was applied to all units of that class. At the completion of this modification the performance had not been helped, but by continuing the tests with graphic instruments one of their engineers in the field solved the problem and eliminated the trouble.

On a recent lot of locomotives with an 80-mile-per-hour gearing it was found that there was a severe sagging in the horsepower curve at 50 to 55 miles per hour whereas the design curve showed this break in the curve should not occur until a speed of 69 miles per hour was reached. As was the case with another make of locomotive, we were told that many similar units were sold to other roads who made no complaints of this nature. The latest advice on this matter is that through some miscalculations in the engineering department traction motor field shunting resistors of improper values were applied.

Invaluable information has been obtained by road testing with graphic instruments in the test car and shortcomings have been detected which never could have been found by load testing devices in the shop. Both types of testing are considered essential on our road.

We have always co-operated with manufacturers in test projects and feel that our findings have been beneficial both to us and to the industry as a whole.

Discussion

R. J. Corfield (Kennecott Copper Corporation, Garfield, Utah): The authors are to be complimented on their fine discourse on operating difficulties pertaining to diesel-electric locomotives and many of their comments and conclusions may be applied aptly to all-electric motive power.

Their conclusions under the heading of switchgear and control equipment are appreciated, particularly this part:

> "Control equipment on some locomotives is installed in such a compacted mass that it is almost impossible to reach some devices or terminals to clear faults which develop. These places are not necessarily inaccessible during the construction at the manufacturing plant, but are rendered so when the hood or housing is placed over them. If design engineers had the responsibility for maintenance, it is felt they would give more thought to the accessibility feature."

It has always been a mystery to me why the manufacturers find it necessary to make such a hodge-podge arrangement of equipment, paying no attention to the customer's particular operating conditions or to their maintenance problems. They should be made to realize that this neglect is detrimental and costly to both manufacturer and customer.

We have stressed this problem for years and it is gratifying to see it brought out in the open.

A Dial-Reading Translator for Digital Machine Inputs

ARTHUR L. KLEIN
NONMEMBER AIEE

KENNETH P. GOW
MEMBER AIEE

Synopsis: At the Southern California Co-operative Wind Tunnel, part of the need for rapid and accurate recording of instrument readings on tabular data sheets has been met by the development of a new automatic-translating device. This device has unusually low torque and is especially suitable for use with self-balancing potentiometers. The device picks up the reading as a whole number and electrically transfers it to standard printing and punched-card machines, without lag and without restricting the normal operation of the potentiometer.

BECAUSE of the high operating cost of a large transonic wind tunnel, such as the Southern California Co-operative Wind Tunnel, at Pasadena, Calif., it is of considerable importance that the experimental data be recorded as quickly and efficiently as possible without sacrificing reliability. The laborious and time-consuming task of permanently recording the many measurements required during the testing of models of modern airplanes has led to the recent development of several automatic recording devices which pick off the reading or position of a force measuring instrument, translate the reading into a whole number, and then print the number on a data sheet. However, to the authors' knowledge, none of these devices could be adapted easily or efficiently to reading electric strain gauges.

Increasing use is being made of small balances, mounted inside the aircraft model, which have strain gauges as the force sensitive element and for which a self-balancing potentiometer is the force-indicating instrument; see Figure 1. For normal tests at the Co-operative Wind Tunnel a battery of three to eight of these potentiometers, representing three to eight forces or moments and requiring the continuous attention of two to five technicians, must be read for each model configuration, model attitude, and Mach number. Clearly it would be advantageous and profitable to employ reliable automatic translators of potentiometer readings which would require the attention of only one technician regardless of the number of channels being read. Such a device has been developed at the Co-operative Wind Tunnel and is described in this paper. It eliminates manual recording from potentiometers of the self-balancing type. The strip-chart type of instrument is not satisfactory because it is less accurate than the dial-indicating type and because its record must be read and manually re-entered on data sheets or punched cards for further calculation.

Design Principles

In June 1947, the senior author undertook the design of a translator suitable to the needs of the Co-operative Wind Tunnel operations. The design principles which he set up were essentially as follows:

1. The translator should pick up the reading of a standard, precision, self-balancing potentiometer and transmit that reading as a whole number, either positive or negative, to a standard punched-card tabulating machine.
2. The translator should follow the movement of the potentiometer without lag even though a printing cycle is taking place.
3. Standard parts should be used as much as possible, wear and maintenance must be minimized, and the dragging torque imposed on the potentiometer must be negligible.
4. The translated reading must be clear and unambiguous.

By December 1947, the basic design work was completed by the senior author and during the following year the details, including the double brush system for tens transfer and the double zero elimination, were for the most part filled in by the junior author. In May 1949, the operation of the basic components was checked out with a temporary gear system, and by November 1949 the completely assembled system was operating satisfactorily.

General Arrangement

The over-all mechanical and electric system for a typical model test is shown

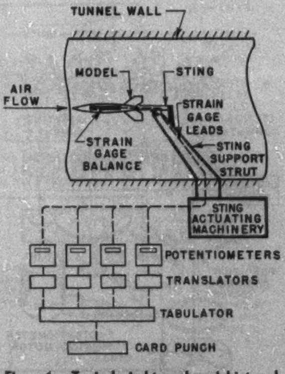


Figure 1. Typical wind tunnel model internal balance installation

---

Paper 52-118, recommended by the AIEE Air Transportation and Computing Devices Committees and approved by the AIEE Technical Program Committee for presentation at the AIEE North Eastern District Meeting, Binghamton, N. Y., April 30–May 2, 1952. Manuscript submitted April 3, 1950; made available for printing February 28, 1952.

ARTHUR L. KLEIN is with the California Institute of Technology, Pasadena, Calif. KENNETH P. GOW was with the California Institute of Technology, Pasadena, Calif., at the time the paper was written, and is now with North American Aviation, Inc., Downey, Calif.

The authors are indebted to members of the staff of the Co-operative Wind Tunnel and of the Guggenheim Aeronautical Laboratory at the California Institute of Technology for their invaluable help and suggestions. J. E. Smith, Assistant Director of the Co-operative Wind Tunnel, made especially important contributions to the present paper.

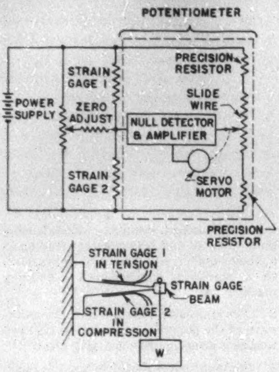


Figure 2. Typical strain gauge and potentiometer circuits

schematically in Figure 1. Figure 2 shows the circuit diagram for a typical strain gauge setup. These systems are quite common and will not be discussed here. Actually the translator interprets only the potentiometer dial reading and does not care what electric signal is fed into the potentiometer. The connection between the two is purely mechanical, and, in fact, the translator is not restricted to use with a self-balancing potentiometer, but may be used to indicate the angular position of any rotatable shaft provided sufficient driving power is supplied by the shaft. It is preferred, but not necessary, that the potentiometer dial be a linear scale. If the scale is nonlinear, however, it is possible to make the proper adjustment in the data reduction process.

Figure 3 is a simplified sketch of the physical arrangement of the translator. A series of spur gears is connected directly to the shaft of the potentiometer servo motor which also turns the potentiometer slidewire and dial to the balance position. Precision-built gears are utilized to reduce the friction torque and minimize maintenance. By the proper selection of bearings and gears and by careful aligument of the gear train, it was found that a backlash removal system was unnecessary. The entire gear train is mounted inside an aluminum alloy casting which is bolted to the chassis of the potentiometer. To the outside of this gear box are mounted four nonrotating commutators, serving as switch contacts, each concentric with a gear shaft as shown in Figure 3. The center of each commutator has been bored out to permit the extension through it of the gear shaft. To these protruding ends of the gear shafts are attached the brush holders and their silver graphite brushes, the brushes riding on the ends of the commutators. Electric power is supplied to one rotating brush on each gear shaft through the end of the shaft which makes direct contact with a flat band of phosphor bronze; this type of rotating contact requires very little driving torque because it operates at such a small radius. For the three commutators which have double brushes, see Figures 3 and 4, power is supplied to the second brush through a conventional slip ring arrangement mounted near the end of the shaft. By using the proper combinations of gear ratios and number of commutator segments it is readily apparent that the position of the rotating brushes can be made to be equivalent to the position of the potentiometer dial.

If we assume, for example, that the entire length of the potentiometer dial is divided into 10,000 equal divisions and that the over-all gear ratio between the dial or slidewire shaft and the units counter gear shaft is 1,000 to 1, then one revolution of the units gear shaft corresponds to a dial travel of ten divisions, and, similarly, one-tenth of a revolution of the units gear shaft is equivalent to a dial travel of one division. Now, if the units commutator has ten equal segments, the movement of the rotating brush in going from one segment to the next represents a movement of one division of the potentiometer dial or a "count of one" change in the dial reading. If each segment is wired to a lamp in such a way that the rotating brush or switch completes the electric circuit through a segment and its lamp when making contact with that segment, then as the dial traverses the units brush rotates and causes each segment light to flash in succession with the cycle being completed once every ten dial counts. In similar manner light circuits can be set up for the tens, hundreds, and thousands commutator counters so that one light in each counter will be lit at a given dial position and the group of four lights will indicate the dial reading to the nearest division.

The physical arrangement just described is quite conventional, being used in automobile mileage indicators in the tenths column, watt-hour meters, and so forth. But such a simple electric circuit could not be used because the reading is not definite; that is, two lamps in the same counter could be energized simultaneously, the units and tens counters would not necessarily reach a count of ten simultaneously, the units, tens, and hundreds counters would not necessarily reach a count of 100 simultaneously, and so forth. In other words, when the potentiometer dial reading is 099 this simple lamp circuit might show such readings as 009, 090, 109, 199, and so forth To ensure that the correct reading

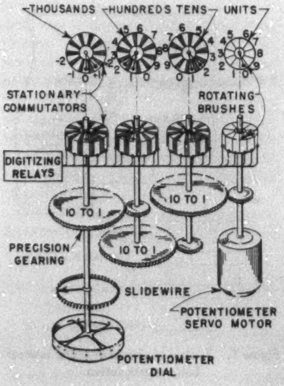


Figure 3 (left). General mechanical arrangement of dial-reading translator

Figure 4 (right). Back side of translator showing commutators and rotating brushes

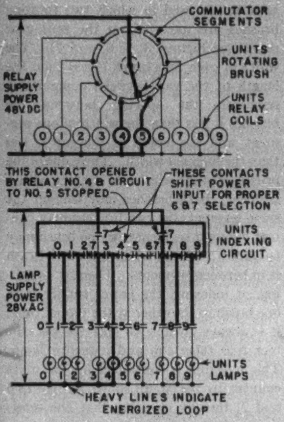

COMMUTATOR SEGMENTS
UNITS ROTATING BRUSH
UNITS RELAY COILS
RELAY SUPPLY POWER 48V.DC
0 1 2 3 4 5 6 7 8 9
THIS CONTACT OPENED BY RELAY NO. 4 & CIRCUIT TO NO. 5 STOPPED
THESE CONTACTS SHIFT POWER INPUT FOR PROPER 6 & 7 SELECTION
UNITS INDEXING CIRCUIT
LAMP SUPPLY POWER 28V.AC
UNITS LAMPS
HEAVY LINES INDICATE ENERGIZED LOOP

Figure 5. Circuit for digitizing units, shown at a count of 4

= represents an open contact unless the relay coil is energized

≠ represents a closed contact unless the relay coil is energized

is given by the lamps, it is necessary to superimpose control circuits upon the lamp circuits which will prevent erroneous lamp readings. The control circuit (considering now the process of the dial reading going from 099 to 100) must hold the hundreds lamp reading at 000 and the tens lamp reading at 90 until the units reading of 9 changes to 0; then simultaneously with the change from 9 to 0 in the units lamps the tens lamp reading must change to 00 and the hundreds lamp reading must change to 100. These control circuits are called indexing systems and are described in the next section.

One other rotating part which is included in the translator is the a-c drag cup induction generator. This generator acts as a damper on the motion of the servo motor and is directly connected to the motor shaft. The generator output, which is manually controllable through a voltage divider, is connected to the potentiometer amplifier and modulates the signal of the amplifier to the servo motor, thereby giving an effective dragging action proportional to the motor speed.

The inertia of the entire translator-potentiometer system is not more than 10 per cent greater than that of the potentiometer in its original condition, and the frictional torque is about the same; note that originally the potentiometer had a worm gear drive. Thus the speed of response and the sensitivity of the potentiometer have not been modified appreciably.

Indexing Methods

With the mechanical arrangement of Figure 3 it is clear that when a rotating brush makes contact with one and only one commutator segment a closed electric circuit can be formed, including an indicator light, which will indicate the number of the commutator segment on which the brush is resting. But when the brush is midway between two adjacent segments the brush will contact both segments, or neither, depending upon whether the width of the brush face is larger or smaller than the gap between the segments. If the brush is wider than the gap both lamps will be lit, while if the brush is smaller than the gap no lamps will be lit. Certainly neither of these conditions is tolerable in our case even though the true correlation of brush position to potentiometer dial reading requires that the brush be midway between the two segments. A choice of numbers must be made either manually or automatically. A manual method would defeat the purpose of the instrument. Therefore, an automatic method must be devised to make the instrument pick one number or the other and do it consistently. This requires a "snap action" or indexing system. The method of indexing probably is the most important feature of any translator.

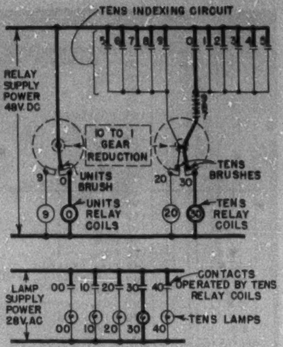

TENS INDEXING CIRCUIT
RELAY SUPPLY POWER 48V.DC
10 TO 1 GEAR REDUCTION
UNITS BRUSH
TENS BRUSHES
UNITS RELAY COILS
TENS RELAY COILS
LAMP SUPPLY POWER 28V.AC
CONTACTS OPERATED BY TENS RELAY COILS
TENS LAMPS
00 10 20 30 40

Figure 7. Circuit for controlling the tens transfer, shown at a count of 30

Mechanical indexing devices, for example, the Geneva motion system used in automobile mileage counters, are subject to rapid wear because of sliding surfaces and impact loads and have rather low speed limits because of the high accelerations encountered. For instance, one can

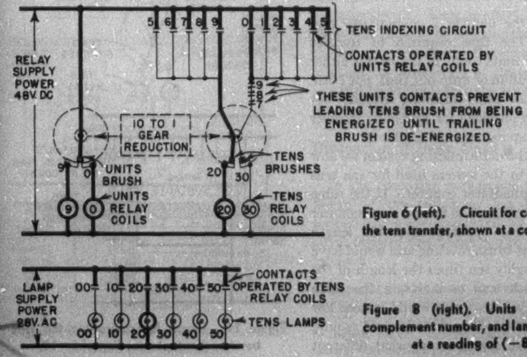

TENS INDEXING CIRCUIT
CONTACTS OPERATED BY UNITS RELAY COILS
THESE UNITS CONTACTS PREVENT LEADING TENS BRUSH FROM BEING ENERGIZED UNTIL TRAILING BRUSH IS DE-ENERGIZED
RELAY SUPPLY POWER 48V. DC
10 TO 1 GEAR REDUCTION
UNITS BRUSH
TENS BRUSHES
UNITS RELAY COILS
TENS RELAY COILS
LAMP SUPPLY POWER 28V.AC
CONTACTS OPERATED BY TENS RELAY COILS
TENS LAMPS
00 10 20 30 40 50

Figure 6 (left). Circuit for controlling the tens transfer, shown at a count of 29

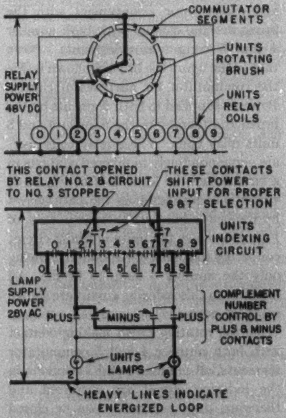

COMMUTATOR SEGMENTS
UNITS ROTATING BRUSH
UNITS RELAY COILS
RELAY SUPPLY POWER 48V DC
THIS CONTACT OPENED BY RELAY NO. 2 & CIRCUIT TO NO. 3 STOPPED
THESE CONTACTS SHIFT POWER INPUT FOR PROPER 6 & 7 SELECTION
UNITS INDEXING CIRCUIT
LAMP SUPPLY POWER 28V AC
PLUS
MINUS
COMPLEMENT NUMBER CONTROL BY PLUS & MINUS CONTACTS
UNITS LAMPS
HEAVY LINES INDICATE ENERGIZED LOOP

Figure 8 (right). Units indexing, complement number, and lamp circuits at a reading of (−8)

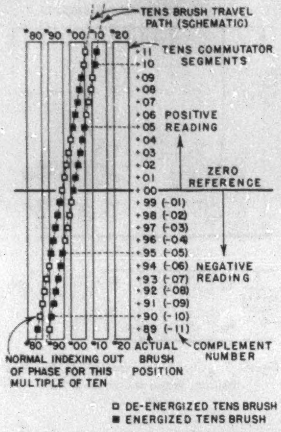

TENS BRUSH TRAVEL PATH (SCHEMATIC)
80 90 00 10 20
TENS COMMUTATOR SEGMENTS
+11
+10
+09
+08
+07
+06
+05
+04
+03
+02
+01
+00
POSITIVE READING
ZERO REFERENCE
+99 (-01)
+98 (-02)
+97 (-03)
+96 (-04)
+95 (-05)
+94 (-06)
+93 (-07)
+92 (-08)
+91 (-09)
+90 (-10)
+89 (-11)
NEGATIVE READING
80 90 00 10 20 ACTUAL BRUSH POSITION
COMPLEMENT NUMBER
NORMAL INDEXING OUT OF PHASE FOR THIS MULTIPLE OF TEN
□ DE-ENERGIZED TENS BRUSH
■ ENERGIZED TENS BRUSH

Figure 9. Schematic layout of tens brush positions for positive and negative readings

visualize easily the difficulties which would be met if the rotating brush were required to jump rapidly from segment to segment so as to prevent contact with two segments at the same time. For these reasons it was decided to delegate the all important indexing action to electric relays, a decision which was supported further by the knowledge that multicontact relays were required to provide additional circuits for interlocking and monitoring functions. The particular circuit chosen is shown in its simplest form in Figure 5.

The upper half of Figure 5 is a schematic end view of the units rotating brush and commutator arrangement with the brush making contact with both number 4 and 5 commutator segments. Note that in the figure the brush is shown inside the commutator for purposes of clarity. Power is supplied to both segments by the brush and thereby to the number 4 and 5 units relay coils. These relay coils operate correspondingly numbered contacts in a circuit (see the lower half of Figure 5) which is energized along the path shown by the heavy lines. The lower portion of Figure 5 shows the units indexing circuit connected to a bank of lamps and allowing only the number 4 lamp to be lit. All contacts in this indexing circuit are shown in their normally de-energized position. With the rotating brush making contact with both number 4 and 5 commutator segments, all number 4 and 5 contacts in the indexing circuit are energized and therefore their normally open or closed positions are reversed. Study of the circuit will show that current can pass through only the number 4 lamp; the circuit to the number 5 lamp is blocked by the combination of the number 4 and 7 contacts. As soon as the rotating brush moves completely off of the number 4 commutator segment going toward the number 5 segment, the number 4 units relay coil is de-energized and all number 4 contacts revert to their deenergized positions, thereby leaving a clear circuit through the number 5 lamp. Similarly, each successive lamp will be lit as the brush revolves through a full revolution with the lower number always taking preference. The indexing loop or circle is always broken by the four special number 7 contacts, with the left-hand power input line being used only when the brush touches the number 7 commutator segment. The time required to transfer from one units count to the next is either a relay pull-in or drop-out time, not both. A parallel units indexing circuit, not shown, is provided for supplying number signals to punched-card machines.

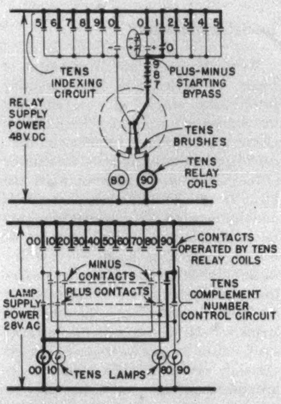

5 6 7 8 9 0 0 1 2 3 4 5
TENS INDEXING CIRCUIT
PLUS-MINUS STARTING BYPASS
9
8
7
RELAY SUPPLY POWER 48V.DC
TENS BRUSHES
TENS RELAY COILS
80
90
00 10 20 30 40 50 60 70 80 90
CONTACTS OPERATED BY TENS RELAY COILS
MINUS CONTACTS
PLUS CONTACTS
LAMP SUPPLY POWER 28V.AC
TENS COMPLEMENT NUMBER CONTROL CIRCUIT
00 10
TENS LAMPS
80 90

Figure 10. Circuit for controlling the tens transfer, shown at a reading of (−9)

From the units indexing system we now progress to the system used for the tens brush-commutator counter. If the same system were used as for the units counter it is readily apparent that the length of the tens-brush rotating arm would have to be roughly ten times the length of the units-brush arm to maintain the same sensitivity. Since this would impose low maximum speed limits and would require much more space, a somewhat different approach is used in which two rotating brushes are mounted on the tens gear shaft (see Figures 3, 4, 6, and 7). The width of each tens brush is less than the length of the gap between commutator segments, the spacing between the brushes is less than the width of one segment, and the length of the gap is less than the segment width. Thus, one of the brushes always will be in contact with a segment; both brushes may contact the same segment, the two brushes may contact different segments, or one brush may contact a segment while the other brush is in between segments. Then by switching, or indexing, the power supply from one brush to the other at the proper time, it is always possible to energize the correct tens relay coil and indicator lamp. This indexing is controlled by the units counter, by the circuit shown in Figures 6 and 7, thereby ensuring that the sensitivity of the tens count is equivalent to that of the units count. The double brush system requires two more brushes (one brush rotating on the commutator and one slip-ring brush for power delivery) than does the single brush system but the rubbing speeds are only one-tenth as large.

The diagrams of Figures 6 and 7 show a simplified layout of the tens brush and indexing system with the circuits set up for counts of 29 and 30 respectively. A portion of the units commutator and relay system is included for reference. Again it should be noted that the brushes are shown inside the commutator for reasons of clarity. Contrary to the units system, the tens indexing circuit controls

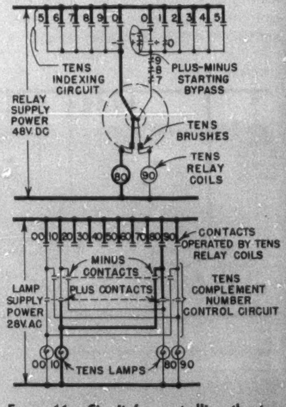

5 6 7 8 9 0 0 1 2 3 4 5
TENS INDEXING CIRCUIT
PLUS-MINUS STARTING BYPASS
9
8
7
RELAY SUPPLY POWER 48V.DC
TENS BRUSHES
TENS RELAY COILS
80
90
00 10 20 30 40 50 60 70 80 90
CONTACTS OPERATED BY TENS RELAY COILS
MINUS CONTACTS
PLUS CONTACTS
LAMP SUPPLY POWER 28V.AC
TENS COMPLEMENT NUMBER CONTROL CIRCUIT
00 10
TENS LAMPS
80 90

Figure 11. Circuit for controlling the tens transfer, shown at a reading of (−10)

the relay power transmission ahead of the rotating brushes and is simpler because it is only necessary to index, that is, switch the power supply from one brush to the other when the units count goes from 9 to 0 or vice versa.

In Figure 6 the units rotating brush is shown contacting units positions number 9 and 0 with the result that units indicator lamp number 9 is energized in the same manner that lamp number 4 was energized in Figure 5. At the position shown in Figure 6, both the number 9 and 0 normally open contacts in the tens indexing circuit are closed, and a number 9 normally closed contact, now open because units relay coil number 9 is energized, is blocking the circuit to the leading (right hand in Figure 6) tens brush so that only the number 20 tens relays coil is energized and the number 20 tens lamp is lit. Transfer, or indexing, in the tens decade occurs, as shown in Figure 7, when the units brush has just broken contact with units position number 9 and is on position number 0. At this position, the units number 9 relay coil is de-energized which opens the only circuit to the trailing tens brush and completes the circuit to the leading tens brush, thus energizing the number 30 commutator segment and relay coil and lighting the number 30 lamp. As indicated in Figures 6 and 7, units counts of 5, 6, 7, 8, and 9 energize the trailing tens brush, and units counts of 0, 1, 2, 3, 4, and 5 energize the leading tens brush. At the number units position, both tens brushes contact the same tens commutator segment and the tens indexing circuit allows both brushes to be energized, thereby increasing the tolerance on brush position in the tens circuit.

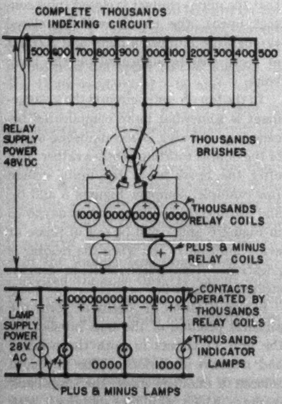


Figure 12. Circuit for controlling the thousands and plus-minus transfer

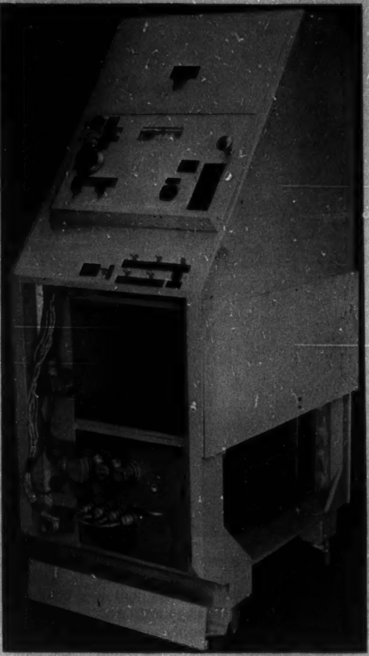


Figure 13. Prototype dial-reading translator set up for operational tests in a 2-channel mockup cabinet

The indexing systems for the hundreds and thousands counters are similar to that used for the tens counter, in that the same type of double brush circuit is employed, with the indexing circuit controlling the energizing of the brushes. The tens relay coils operate the contacts in the hundreds indexing circuit, and the hundreds relay coils operate the contacts in the thousands indexing circuit. Thus, in effect, the units relay coils control the indexing throughout the entire system and by that means establish that the sensitivity of the entire system will be equivalent to the sensitivity of the units counter.

Discussion of Basic Design

In the preceding paragraphs the basic design of the translator has been described in considerable detail. The major advantages of this design can be stated as follows:

1. Speeds greater than the maximum continuous counting speed (relay response time) do not cause damage. The maximum rate of brush travel, in the units counter, is nearly 600 counts per second, which figure is determined by the desired number of counts across the potentiometer scale and the maximum rpm, 1,800, of the potentiometer servo motor. The relays follow at speeds up to about 50 counts per second.

2. The correct count or reading is obtained automatically after a high speed, greater than 50 counts per second, run when the balance position is reached, irrespective of the number of counts missed by the relays. This means that the translator does not restrict the motion of the potentiometer and that it does not have to count every number when traversing from one position to another.

3. The dragging torque of the translator is independent of the potentiometer dial position.

4. The system counts down as easily as up.

5. Indexing, which requires snap action by some mechanical part, is handled by standard relays and contactors which have had many years of development to achieve reliability in this type of operation.

6. If the translator power is interrupted for any length of time the correct count is immediately re-established when the power is restored.

7. It is believed that the maintenance requirements of the translator will be small.

8. The translator may be adapted easily to many other data producing systems.

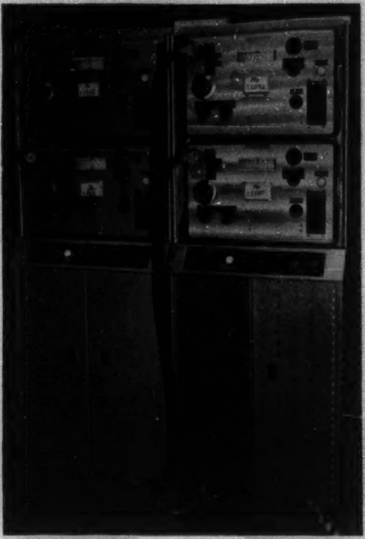


Figure 14 (left). Two chassis containing 4 channels of the production version of the dial-reading translator

Figure 15 (right). View of the chassis with the doors and sides removed, showing the relay chassis, power supply, and a Leeds and Northrup amplifier for the lower channel

In constructing the prototype unit it was found more convenient to use a gear reduction of 1 to 5 between the units and tens shafts than the 1 to 10 ratio shown in the sketches, and the number of units counts or segments on the units commutator was increased to 20, that is, the units brush goes from 0 to 9 twice in one revolution.

The thousands counter is used primarily to control the transfer between positive and negative readings (see discussion in next section) rather than to give readings to four figures. The accuracy of the potentiometer itself is not equal to four significant figures, but, for the sake of clarity in reading, the full dial range is divided into approximately 2,100 counts.

The diagrams in this paper show only the circuits for the count-indicator lamps. In the prototype device a parallel circuit is provided in every case to furnish the count signal to International Business Machines card-punching and printing machines. Still other parallel circuits could be added if desired.

Method for Indicating Negative Numbers

In a normal wind tunnel test the strain gauge will indicate both positive and negative forces, so that it is desirable to read and record the data with the proper reference zero. Circuits have been included in the translator which enable it to indicate negative as easily as positive numbers. These circuits are described in the following paragraphs. It should be noted that Figures 5, 6, and 7 show only those circuits required for positive number indication.

First, assume that the potentiometer is zeroed at the center of its dial and that the translator counters are set to read zero simultaneously. Then if the strain gauge input is increased in the positive direction the translator lamp bank will show in succession the numbers: 0000, 0001, 0002, 0003, and so forth. But if the strain gauge output increases negatively, the potentiometer dial moves in the opposite direction and the translator lamp bank (assuming the negative number circuits have been omitted) will show in succession the numbers: 0000, 9999, 9998, 9997, and so forth. Thus it is necessary to add additional controls which will convert these latter numbers to their complements and also will give a minus sign indication. It is clear that if every individual lamp circuit is connected with its complement (for example, number 9 with 1, 8 with 2, 60 with 40, 400 with 600, and so forth) the addition of plus and minus contacts in these branch leads will give control of the choice between a number and its complement. The arrangement for the units count is straightforward and is shown in Figure 8. A pair of plus and minus normally open contacts is inserted in each of the number 1, 2, 3, 4, 6, 7, 8, and 9 lamp circuits (note that the complements of 0 and 5 are 0 and 5). Only the circuits for the number 2 and 8 lamps are given in Figure 8. When the reading is positive all plus contacts are closed, and when negative, all minus contacts are closed. Thus if the units brush is at the number 2 commutator segment but the reading is actually negative so that the minus contact is closed (plus contact open), the circuit will be closed through the number 8 lamp and a count of (−8) will be indicated on the lamp bank.

In the tens-counter circuits the arrangement is somewhat more complicated because the indexing is accomplished ahead of the double rotating brushes rather than after the brush as in the units system (for example, compare Figures 5 and 6), and because the indexing system as shown in Figure 6 and 7 will be out of phase at every multiple of ten when the reading is negative. This phase shift is demonstrated in Figure 9.

Figure 9 shows in schematic form the physical position of the tens brushes for the first 11 counts in both the positive and negative directions. The left-hand column of numbers shows the lamp bank readings which would be obtained with the circuits of Figures 6 and 7. The right-hand column of numbers, negative half only, shows the correct readings which are desired. Inspection of this array will show that paralleling the number 90 and 00 tens-lamp circuits and inserting plus and minus contacts will give the correct complement number reading (assuming, of course, that the units system of Figure 8 is in operation) for the first nine counts but not for the tenth count. Similarly, paralleling the number 80 and 10 tens-lamp circuits gives the correct complement for the counts of 89 (−11) to 81 (−19) inclusive, but not for the 20th count. Thus we see that at every tenth negative count the device of paralleling the lamp circuits is inadequate in that the entire system is out of phase by one count. The proper phase relationship could be obtained if an extra count (−00) were added between +00 and −01, but this is undesirable because it puts an error in the numbering system and coarsens the zero setting. The method finally adopted was to set up two (0) circuits in the tens indexing circuit with one controlled by a plus contact and the other controlled by a minus contact. The plus (0) circuit is the same as that shown in Figures 6 and 7 and is used only for positive numbers and the reference zero. The minus (0) circuit is used only for negative numbers and has the effect of causing the tens system to index between the 1 and 0 units counts instead of between the 0 and 9 units counts, which effect is the same as the desired phase shift of one count. In Figure 9 this shift would result in the left-hand brush being energized at the number 90 brush position instead of the right-hand brush and, therefore, the number 80 commutator segment being energized instead of the number 90 segment, thus enabling the correct reading of −10 to show on the lamp bank. The tens indexing circuit with the positive-negative controls added is shown in the upper half of Figures 10 and 11 for readings of −09 and −10.

In the lower half of Figures 10 and 11 are shown the paralleled tens-lamp circuits with the plus and minus control contacts. Again only part of the lamp circuits is shown to avoid confusion. Note that the normally closed (0) contact in the (1) circuit of the tens indexing system is required (for negative readings) to prevent the leading tens brush from being energized through this (1) circuit when the units brush is contacting both the number 0 and 1 units commutator segments.

If a diagram similar to Figure 9 is drawn up for the hundreds and thousands counters it will show that the count does not get out of phase as did the unmodified tens counter. This means that the corrective plus-minus circuits of the tens indexing system should not be added to the indexing systems of the hundreds and thousands counters. In fact, the latter must be analogous to the simple tens indexing circuits shown in Figures 6 and 7. However, the complement number circuits in the lamp bank system are similar in all three double-brush counters (see Figures 10 and 11).

To complete this discussion of the control of positive and negative numbers, we have only to mention the method used to actuate the plus and minus contacts; that is, the means used to determine the algebraic sign of the potentiometer reading. Of all four counters only the thousands counter does not rotate more than 180 degrees in either direction; and therefore only in the thousands counter will a commutator segment always be positive or negative. In Figure 12, the circuit arrangement for the thousands counter is shown. When the reading is positive one of the two right-hand commutator segments is energized and a plus signal is given; and when the reading is negative one of the two left-hand segments is energized and a minus signal is given. These plus and minus relay coils control all plus and minus contacts in the translator. If all counters are at zero the leading (right-hand) thousands brush is energized and a plus signal is given; then if the units counter moves negative by one count to the (−1) position the trailing (left-hand) thousands brush is energized, the minus relay coil is energized, and units, tens, and hundreds circuits switch to their complement number systems.

One other characteristic of these systems is that if the relay power supply is turned on with the units or tens counters at their zero position the plus-minus contacts will block the signal to the higher counters and the plus-minus relay coils will not be energized. For example, if the power is turned on with the tens counter at (00) the thousands brush and plus-minus coils will not be energized because the plus-minus contacts in the tens indexing circuit block the circuit to the thousands brush. To remove this difficulty a plus-minus, starting by-pass circuit (see Figures 10 and 11) is included in the tens indexing circuit, thereby allowing the correct plus or minus relay coil and contacts to act, after which the by-pass takes itself out of action.

Automatic Averaging of Fluctuating Balance Loads

One important characteristic of strain gauge signals is the nearly complete absence of any damping in the gauge. If, as is usually the case, the strain-gauge, model-force balance also has no internal damping, then the strain gauge signal coming to the potentiometer frequently will oscillate rapidly about the mean value. Two special devices are included in the translator-potentiometer combination which enable the wind tunnel operator to obtain an average reading.

One device is the 60-cycle a-c drag cup induction generator which is mounted in the translator gear box and whose armature is driven by the potentiometer servo motor with a 1 to 1 speed ratio. The generator output, which is directly proportional to the revolutions per minute of the servo motor, is put into the potentiometer amplifier in such a way that the strength of the amplifier signal to the servo motor is reduced. By this means the response of the motor is reduced and its ability to follow the strain gauge oscillations is curtailed. Enough damping action is provided so that with full damping the potentiometer metrical circuit unbalance required to produce full servo motor speed is several times the unbalance required with damping.

The other device is a low-pass filter in the strain gauge signal input. All force oscillations with frequencies above 1/2 cycle per second can be attenuated and in most tests effectively eliminated. Several filters are provided, each with different cutoff frequencies, and various combinations of them may be selected as desired. The filter circuit does not modify the speed of response of the servo motor.

These two devices have been in operation in routine wind tunnel tests and have proved to be quite satisfactory as well as necessary.

Additional Features

Several other features of the translator are described briefly now.

The indicator lamps are combined in a 4-column bank of the instrument panel (see Figure 13) and light up numbered jewels. Provisions for operating a remote lamp bank are provided.

Data printing circuits are paralleled with the indicator lamp circuits and connected to International Business Machines standard punch-card machines. The printing cycle is controlled by a push button. Plus and minus signs are shown on the lamp banks and recorded in the printing cycle with the numerical reading. When the printing cycle is initiated all commutator relays are locked up to hold the reading while the brushes continue to rotate and follow the potentiometer reading, that is, the printing operation does not interfere in any way with the normal action of the potentiometer. The lamp bank reading is also "locked up" during the printing cycle. Additional circuits are provided which prevent printing a reading in the event that there is no number or more than one number in any of the four decades; this condition may exist momentarily during high-speed travel of the potentiometer dial and translator, or because of faults in the relays.

The potentiometer scale zero reference may be placed at the center of the dial or displaced to fixed positions at either end of the dial. This amounts to adding a "pan weight" to the reading. The translator thousands counter has a special circuit which causes the translator to indicate the correct reading regardless of which zero reference is used.

Conclusion

Figures 14 and 15 show two views of the 4-channel chassis.

The prototype translator, shown in Figure 13, has been used to record data in many routine wind tunnel tests and has operated quite satisfactorily. The translator cabinet shown in Figure 13 is a wooden mockup and is to be replaced by a steel frame. The Co-operative Wind Tunnel is constructing enough additional units to enable it to record automatically all strain gauge balance data. It is believed that this device will prove to be a very useful addition to wind tunnel instrumentation.

Discussion

Harry E. Burke (Consolidated Engineering Corporation, Pasadena, Calif.): I find this paper very interesting, particularly because I have had the opportunity to examine the equipment as it was operated in conjunction with the Co-operative Wind Tunnel. As far as I have been able to ascertain, the unit described is entirely satisfactory in this application. The method of damping out cyclical variations imposed on the input signal is straight-forward and is a very useful feature. The required toggle action in stepping from one digit to the next is accomplished in a relay matrix which also is used to store a particular digital reading while the next reading is being set up on an analogue basis. This also has some very important advantages in some applications.

The only criticisms that I can make are relatively minor ones and are not intended as reasons for curtailing the use of this equipment. The use of an analogue section and a digital section in this one instrument appears to be rather expensive and it is felt that if these functions could be combined together in some way then the cost of digitizing the input signal would be reduced greatly.

It is my belief that the equipment as it is described would be capable of an accuracy and a linearity of about ±1 part out of a possible full scale of 1,000 parts. As this accuracy depends in a great part on the slide-wire used in the analogue section and in the gear train driving the digital section, it seems to me that there would be a deterioration in these specifications as the slide-wire and the gear train wear with use.

I found the demonstration of this unit as a standard part of the Co-operative Wind Tunnel's instrumentation very interesting and I would like to congratulate the authors of this paper for its excellent design.

Arthur L. Klein and **Kenneth P. Gow**: Mr. Burke's statement in his second paragraph is quite correct. The instrument would be much simpler without the potentiometer and its accessories. These particular instruments were designed to be used with strain-gauge force measuring equipment. A simple digitizing unit would require only the gear box and, without special features, only four relays. The inaccuracies in the instrument lie almost entirely in the potentiometer. The gear train commutators and their equipment have intrinsic accuracies of at least two orders of magnitude higher. The writers believe that with proper precaution a gear box system, as described, can be made with an error of less than 0.0001 per cent.

Unit Equipments for Aircraft Electric Power Systems

H. J. FINISON
MEMBER AIEE

THE rapidly expanding use of electric apparatus on aircraft since 1940 has given emphasis to the development of new devices and machines of increased rating and reduced weight for aircraft electric power systems. Outstanding progress has been made in improved designs of reduced weight and space and increased performance and life. Much work still remains in device engineering to work continually toward further improvements as new materials and new techniques become available. Increase in aircraft operating demands will probably continue to mean electric power systems of increased size and improved reliability. These requirements will call for more work in development of devices of increased rating and performance. Still further, many of the devices currently used have not had the benefit of the improved design techniques and materials that are evident in some of the more advanced designs. Hence, it appears that engineering problems related to designs of individual devices—device engineering—will continue to be an important part of engineering for aircraft electric systems.

System Engineering

As the work on improved devices continued, and the size and complexity of the electric power system increased, it became evident that improved devices in themselves could not guarantee successful electric system performance. Thus aircraft electrical system engineers began to be concerned with the performance of these devices in an integrated system. Some of our early concerns in this field dealt with problems of contact failure on contactors. Aircraft electrical system engineers found that on some systems voltage would drop so low that even the best of contactor designs could not hope to remain closed and avoid contact welding during motor starting. In fact, on one airplane it was found that during reversal of a large motor drive system, voltage was not only low but went through zero and reversed for a short time.[1] It was evident that improvement in the contactors as devices could never be successful in solving this problem, but that over-all system engineering was essential to study the relationships between power supply system and load demands.

Aircraft electrical engineers have been active on system engineering problems on aircraft for several years, but here too, much work remains to be done. In fact, there is probably considerable need for expanded emphasis in system problems not only in the power supply system but in the entire distribution system in the airplane.

There is still another field of engineering in addition to device and system engineering that is essential to achievement of electric power systems of minimum weight, maximum reliability, reduced maintenance, and cost. This additional field is equipment engineering.

Equipment Engineering

Work in this field involves development of over-all equipments to fit particular unit functions. For example, we know that the controls for a motor must be comprised of a connection to a power bus, a circuit protective device, and one or more contactors, depending on whether the control is for full voltage, reduced voltage, reversing, or nonreversing service. Hence, it is possible to consider the motor control function as a unit function, and assemble the required component devices into a single equipment. Since it is known in advance how the individual devices will fit into their final assembly, it is possible to design those devices to meet exact needs and in the process achieve reduced weight and size.

Here again some work in equipment engineering has been previously carried out. And considerable expansion of effort in this field appears desirable.

Example

To further illustrate the concept of unit equipments, one equipment as designed for generator control service will be described.

Controls for a d-c generator include:

1. Voltage regulator.
2. Switching device for normal switching.
3. Sensing relays for the switching device, including reverse current, differential voltage, and so forth.
4. Circuit protective devices.
5. Additional protective devices for control circuits and overvoltage.
6. Ammeters and voltmeter.
7. Miscellaneous resistors and switches, some of which may be a part of these devices.

Considerable work has been carried out in the past few years in studying the system arrangement for the generator controls, and various systems have been devised. However, once a system is adopted a definite known complement of devices will always be required for each generator.

Illustrated in Figures 1 and 2 is one method of assembling the individual devices into an over-all power center equipment.[2] Figure 3 shows one concept of the location of this power center in the aircraft.

The particular assembly shown possesses a number of advantages:

1. It integrates the controls into a single equipment.
2. The equipment is draw-out.
3. The interconnecting wiring is done at the factory with minimum expense and maximum reliability.

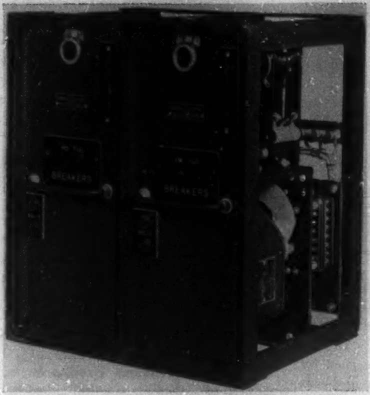
The control equipment shown in Figures 1 and 2 was necessarily designed around existing devices. The shapes of the devices are not necessarily most advantageous to fit into the final assembly. Terminals on devices are not necessarily in the most advantageous places to fit into the over-all assembly. It appears evident that a considerable improvement in the over-all equipment could be achieved by redesign of the devices to fit the final assembly. The redesign is possible since the final physical character of the devices can be established through superior knowledge of their actual location and final use. Some of these improvements have been achieved in later designs of the generator control equipments.

The power center shown by Figures 1 and 2 does not completely meet the objectives of unit equipments. It would be desirable for the assembly to be fully en-

---

Paper 52-135, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE South West District Meeting, St. Louis, Mo., April 15–17, 1952. Manuscript submitted July 27, 1950; made available for printing March 13, 1952.

H. J. FINISON is with the Armour Research Foundation of the Illinois Institute of Technology, Chicago, Ill.



Photograph above courtesy the General Electric Company

Figure 1 (above). Power center incorporating complete controls for two generators in a draw-out framework

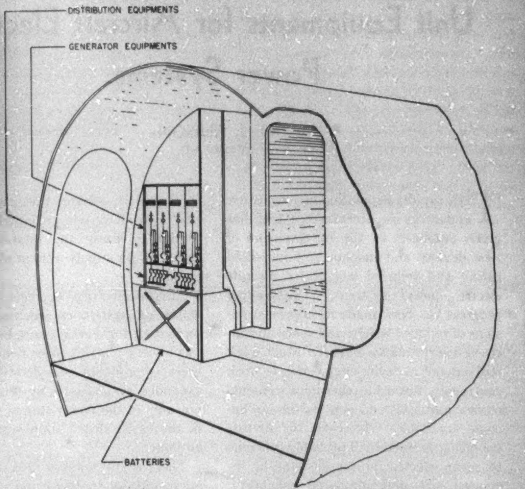


Figure 3 (right). Main power center in a typical transport-type aircraft

closed to eliminate the necessity for the enclosing of individual devices. And, by having the assembly fully enclosed it serves as a protective housing during the time the equipment is in the supply warehouse and while it is being installed in the airplane. One difficulty of fully enclosing the unit involves cooling of the voltage regulator. However, where the voltage regulator is designed specifically to fit the enclosure, it may be designed to allow heat to be conducted to the enclosure and thence radiated or convected to the surroundings. Since a larger cooling surface is thereby available, the voltage regulator cooling problem actually may be alleviated.

As originally conceived, the generator control center would have been complemented by additional equipments, one of which would have been a supervisory center incorporating the ammeter, control switches, and voltmeter, and intended for location at the pilot's or flight engineer's station; see Figure 4. Further, it was originally expected that the power center would either incorporate the main distribution bus or be integrated with such a bus. And protective devices for outgoing feeder circuits would be incorporated into the over-all power center equipment; see Figures 3 and 5.

Experience in Other Fields

While the concept of unit equipments is relatively new to aircraft systems, there is considerable background of experience in other fields. For example, since 1938 industrial and public utility power systems have gone through a complete revolution in the arrangement of their systems. Formerly, individual component devices for a substation, for example, were procured as individual devices, shipped to the job site, and assembled piece by piece on the job. Since 1938 metal-clad equipments have been developed which permit complete assembly of the devices in the factory; see Figure 6. By redesign-

Photograph below courtesy the General Electric Company

Figure 2 (below). Unit equipment for generator removed from power center shown in Figure 1

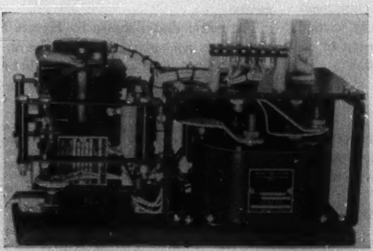
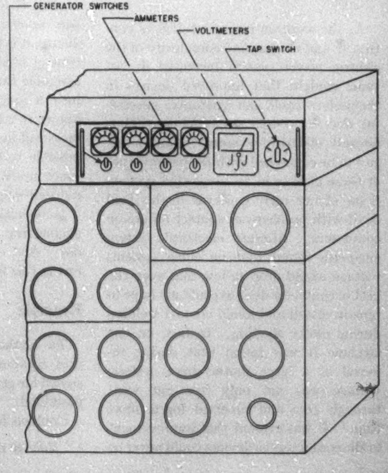


Figure 4 (right). Supervisory equipment for aircraft d-c generator system



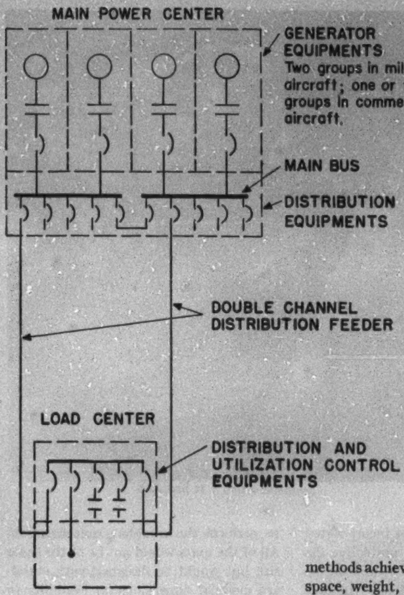

MAIN POWER CENTER
GENERATOR EQUIPMENTS
Two groups in military aircraft; one or two groups in commercial aircraft.
MAIN BUS
DISTRIBUTION EQUIPMENTS
DOUBLE CHANNEL DISTRIBUTION FEEDER
LOAD CENTER
DISTRIBUTION AND UTILIZATION CONTROL EQUIPMENTS

Figure 5 (left). One-line diagram illustrating use of unit equipments to form a power center and a load center

Courtesy the Melpar Company

Figure 7 (above). Potted assemblies are being used for electronics equipments to reduce space, weight, cost, and to improve maintenance

ing of the individual components to fit the final assembly and by extensive engineering applied to the problems of assembling, outstanding reduction in space, cost, and engineering required for each job have been possible. Work with such metal-enclosed unit equipments has extended into motor control—low voltage distribution equipment and transformer and switchgear combinations.

Similarly, considerable work is now being done on development of improved assembly methods in the electronics field, including work with printed and stamped circuits and potted assemblies; see Figures 7 and 8. Here the improved methods achieve considerable reduction in space, weight, and cost.

Present Methods

To illustrate further the type of problems we are trying to solve with the unit equipment concept, two typical aircraft installations will be reviewed. Figures 9 and 10 show junction boxes on one airplane. Each houses a load bus, protective devices, and control components for a number of circuits. In general, such installations are made up from standard size junction boxes in which is mounted an insulating board, the copper bus protective devices, and control devices of various shapes and sizes. The individual devices are interconnected with specially built cables or copper bars as required. Usually incoming and outgoing cables terminate on the individual devices.

Critical examination of these installations reveals the following deficiencies:

1. Considerable waste space—the individual devices are not designed to fit well together.
2. Considerable weight in interconnecting bus and cables—terminals of devices are not located to keep interconnections to a minimum length.
3. The assembly must be engineered for a particular airplane. Changes in electric load may mean complete redesign of the entire assembly.
4. Considerable labor is required for assembling and in making the individual connections.

Principles for Aircraft Unit Equipments

A few principles which should guide the development of equipments are:

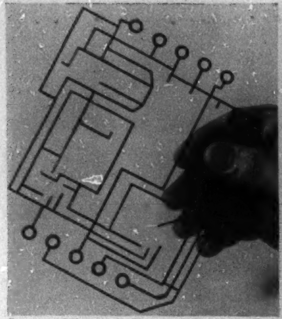
1. The device housing should be the final housing of the equipment—no box inside a box inside a box, and so forth.
2. Where possible, the mechanical connections for installing the equipment should also provide the electric connections.
3. Use of a building block scheme with standard module dimensions so that individual units will fit into almost any system arrangement.

Illustrative Example—Load Center Equipments

To illustrate further the concept of unit equipments, one possible approach

Figure 6. Unit equipments are commonly used in industrial and electric utility power systems as illustrated by the power plant illustration. Directly in back of the generator are seven units of metal-clad switchgear with removable vertical-lift circuit breakers. The unit at the right supplies a transformer which feeds the low-voltage switchgear at the extreme right. These units use draw-out air circuit breakers. At the far left is a group of motor-control equipments with the supply bus enclosed at the top feeding several motor controllers of differing sizes

Courtesy the General Electric Company



Courtesy Minneapolis-Honeywell Regulator Company

Figure 8 (above). Interconnections for an electronic equipment stamped from a single piece of metal

100
BATTERY SOLENOID
CAMERA RELAY
FUSES

Photograph right courtesy United States Air Force

Figure 9 (right). Electric load bus and junction box on one airplane

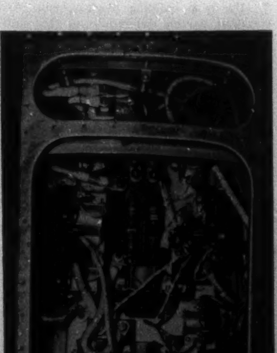
will be given. Although the illustration will appear concrete and definite and some study has been given to the problem in arriving at the illustrative solution, the investigation is not nearly extensive enough to say that this is the only approach to the problem.

Figure 11 shows the elementary diagram of two typical load units. On the left is a full voltage starter unit which incorporates the circuit protective device and the motor contactor. The reduced voltage starter on the right incorporates, in addition to the circuit protective device, the switching contactors and the associated resistors and controls. In each case the combination of devices represents a complete control function, and since the function must always be complete for each installation, the devices can be incorporated into single-unit equipments capable of performing the complete motor control function.

Figure 12 shows one possible physical construction for a typical unit equipment, in this case a full voltage motor control unit. This unit is so designed that it will lock into the next adjacent unit equipments to form a complete load center. The mounting screws also provide the power connections, and control connections are made through plug-type connectors. In the case of the unit illustrated, the equipment would enclose the protective circuit breaker and an individual contactor. Other units would incorporate additional devices as required to perform the complete unit function. All of the units would not be of the same size but would be designed with standard modular dimensions for flexibility in making up a complete load center.

The application of unit equipments can be illustrated by taking a particular load bus on an airplane as shown in Figure 13. Connections to the bus include two incoming lines with assorted circuit protective devices, two large nonreversing motor circuits, three small nonreversing motor circuits, two small reversing motor circuits, one very large nonreversing motor circuit, and one large reversing motor circuit.

Figure 14 shows a possible arrangement for the individual equipments to meet the functions as shown by the Figure 13 diagram. Each of the equipments of which the load center is built may be, with rare exceptions, a completely standardized unit. The equipments and bus assembly should be such that standard connection arrangements are possible



Photograph left courtesy United States Air Force

Figure 10 (left). Power bus and junction box on one airplane

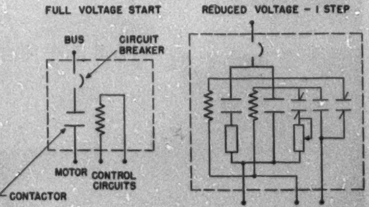


Figure 11 (right). Elementary diagram of typical load control units—full voltage and reduced voltage motor controllers

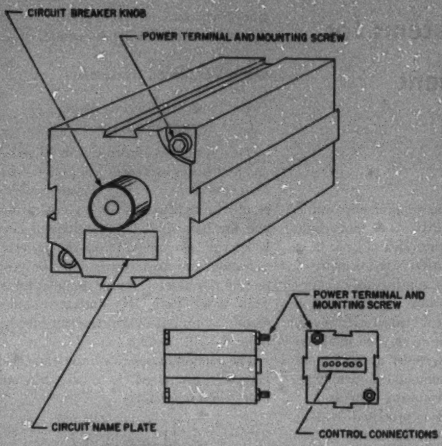

CIRCUIT BREAKER KNOB
POWER TERMINAL AND MOUNTING SCREW
POWER TERMINAL AND MOUNTING SCREW
CIRCUIT NAME PLATE
CONTROL CONNECTIONS

Figure 12. Illustrative example of a full voltage motor controller designed as a unit equipment

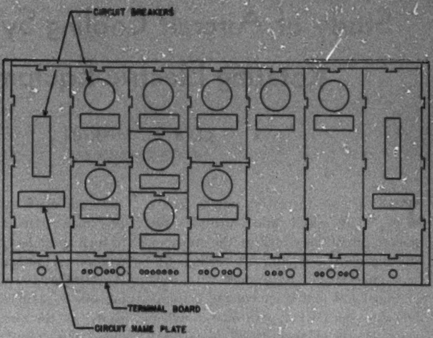

CIRCUIT BREAKERS
TERMINAL BOARD
CIRCUIT NAME PLATE

Figure 14. Front view of load center using unit equipments

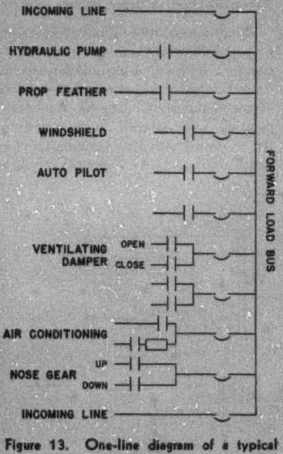

INCOMING LINE
HYDRAULIC PUMP
PROP FEATHER
WINDSHIELD
AUTO PILOT
VENTILATING DAMPER
OPEN
CLOSE
AIR CONDITIONING
NOSE GEAR
UP
DOWN
INCOMING LINE
FORWARD LOAD BUS

Figure 13. One-line diagram of a typical load bus

with connections brought out to the terminal strip for ease in connecting to outgoing cable circuits.

The service replacement of an individual unit can be accomplished by loosening two screws and withdrawing the unit.

Additional Needs

There are many needs for unit equipments for the aircraft electric power system. They include:

1. Generator controls.
2. Generator supervisory equipments.
3. Power center equipments.
4. Distribution (load center) equipments.
5. Utilization control equipments.

Summary of Unit Equipment Concept

The development of unit equipments and the redesign of component devices to fit those equipments can achieve important benefits to improved aircraft performance. It is believed that work in this field can achieve marked improvements by:

1. Reduction in weight.
2. Reduction in space requirements.
3. Improved ease and speed of project engineering for a particular airplane.
4. Improved ease and speed of ordering of component devices.
5. Reduction in airplane manufacturing cost.
6. Improved ease and reduced cost of supply and stocking of replacement parts.
7. Reduction in maintenance time—in elimination of difficulties, ease of understanding the component circuits, and reduction in time for replacement of equipment.

References

1. DESIGN PRINCIPLES FOR ELECTRICAL LANDING GEAR CONTROL SYSTEMS, H. J. Finison. *Aeronautical Engineering Review* (New York, N. Y.), June 1945, pages 5–19.

2. D-C AIRCRAFT GENERATOR POWER CENTER, E. L. Bottemiller. *AIEE Transactions*, volume 66, 1947, pages 1267–70.

No Discussion

Study of Aircraft Cooling Systems for Rotating Electric Equipment

CECIL G. MARTIN
ASSOCIATE MEMBER AIEE

THE steadily increasing upper limits in flight speed and altitude for modern aircraft are making the study of rotating electric equipment cooling systems imperative. Experience with conventional fan and blast cooling systems indicates poor cooling performance under many of the high-altitude, high-speed flight conditions now being encountered. Further, there are strong indications that other types of cooling may extract a lower over-all penalty from the airplane than the conventional systems.

Several generator cooling systems have been compared on a fuel consumption basis by Naumann[1] and an excellent treatment of cooling systems for aircraft electronic equipment, some of which are applicable to rotating machinery, has been made by Robinson and Zimmerman.[2]

There are many criteria which can and must be used in comparing cooling systems. These include:

1. System weight.
2. System size.
3. Ease of installation and maintenance.
4. Range of applicability.
5. Penalty on the aircraft in terms of fuel consumption, and loss of carrying capacity.
6. Reliability and simplicity.

Unfortunately, some of these do not yield to analytical attack and their relative importance varies tremendously with respect to the type of application, time of operation, life expectancy, and environment.

There are almost an infinite number of cooling system variations which could be applied to aircraft. Of course, some would be inherently better than others and each most certainly would have a specific range of applicability. To attempt evaluation of all possible systems under the full range of flight speed, altitude, and installation requirements would be somewhat foolhardy. The magnitude of such an undertaking, lack of basic heat transfer and installation data for many of the systems, and the continuously changing power plant picture prohibit such a study at the present time. However, qualitative comparison may be made to determine which systems appear most feasible; then more detailed study may be made on the most promising systems.

Most of the systems proposed to date may be grouped somewhat loosely as follows:

1. Open ventilated.
2. Totally enclosed.
3. Compressor bleed air.
4. Closed recirculating.
5. Conventional with higher temperature limits.

Each of these system groups properly contains many variations. Diagrammatic sketches of those considered in this paper are shown in Figures 1 through 4. Preliminary investigations and qualitative discussions for each system are indicated in the following discussion.

Open Ventilated Systems

Open Frame Utilizing Free Convention and Radiation

This system is illustrated in Figure 1 (A) and is not ordinarily suitable for aircraft due to poor cooling efficiency, which results in high weight for a given rating, and fire hazard. The quantities of heat which can be dissipated from the frame by radiation and free convection are very limited for normal frame temperatures. This point may be illustrated quite readily by simplified heat transfer calculations.

The maximum possible frame radiation may be computed from the Stefan-Boltzmann relation

$$Q=3.54\times10^{-4}\left[\left(\frac{T_f}{100}\right)^4-\left(\frac{T_a}{100}\right)^4\right] \qquad (1)$$

where

Q = watts per square inch
T_f = frame temperature, degrees Rankine
T_a = ambient temperature, degrees Rankine

Using this equation, Figure 5 has been prepared, showing maximum frame radiation as a function of altitude for several frame temperatures between 100 and 250 degrees Fahrenheit. The ambient temperature was assumed to be equal to the National Advisory Committee for Aeronautics (NACA) standard[3] for each

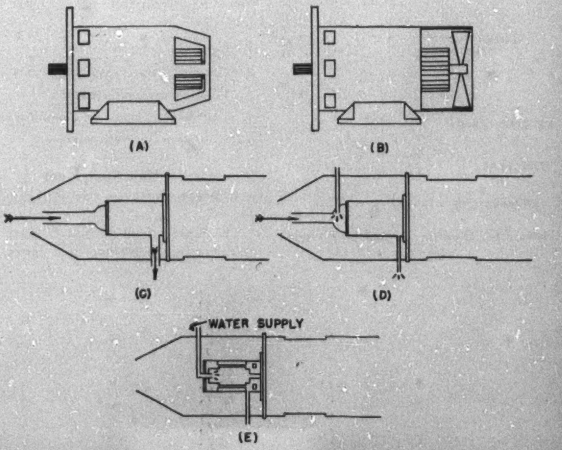


Figure 1. Open ventilated machines. (A) Conventional open machine; (B) Fan-cooled machine; (C) Blast-cooled, jet engine installation; (D) Blast-cooled with water injection, jet engine installation; (E) Evaporative-cooled jet engine installation

---

Paper 52-136, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE South West District Meeting, St. Louis, Mo., April 15–17, 1952. Manuscript submitted December 5, 1951; made available for printing February 27, 1952.

Cecil G. Martin is with Jack and Heintz, Inc., Cleveland, Ohio.

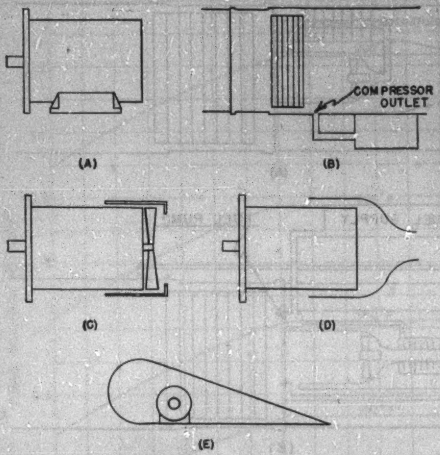


Figure 2. Totally enclosed machines. (A) Totally enclosed; (B) Totally enclosed, pressurized; (C) Totally enclosed, fan air over; (D) Totally enclosed, blast air over; (E) Totally enclosed, heat transfer by conduction to frame

altitude. It will be noted radiation increases with increasing altitude up to the tropopause, then remains constant due to the constant ambient temperature at higher altitudes. Further, the rate of dissipation is quite small considering the losses of present high-performance low-weight small-area machines. The free convection from the frame may be computed approximately from the relation recommended by Brown and Marco[4] for horizontal pipes

$$Q=4.57\times10^{-4}(T_f-T_a)^{1.25}(D)^{-0.25}\left(\frac{P}{14.7}\right)^{0.5} \qquad (2)$$

where

P = pressure, pounds per square inch absolute

D = machine diameter, feet

Again using NACA ambient temperatures, the curves of Figure 6 were computed showing frame free convection as a function of altitude for the same range of frame temperatures. An 8-inch diameter frame was assumed although, as may be noted in equation 2, the resulting heat transfer per unit area would not be affected appreciably by slightly larger or smaller diameters. As can be seen on the curves, the heat dissipation by free convection from the frame is generally much less than that obtainable by radiation, becoming almost negligible at high altitude. In any given installation, of course, air currents across the machine might increase the convection heat transfer appreciably. This, in many cases, would be offset by ambient temperatures higher than the NACA standard values assumed.

The radiation of Figure 5 and the free convection of Figure 6 may be added to give the net heat transfer shown in Figure 7. As indicated by the curves, net frame heat transfer by radiation and free convection are not affected appreciably by changes in altitude. This is certainly a desirable characteristic but the magnitudes are so small as to make this mode of heat transfer impractical for any but small, low output, or intermittent duty machines.

Additional heat may be transferred by conduction from the electric equipment frame to the aircraft frame but this mode will be discussed in a later section.

The radiation and convection per unit area for the rotor of this type machine may be two to ten times that obtainable from the frame depending upon design. The convection heat transfer is often higher than would be expected when the rotor is designed, often inadvertently, to act as a fan. The portion of the transferred heat due to this falls off considerably at altitude as will be shown in the section on fan cooling.

FAN COOLING AND BLAST COOLING

Figures 1(B) and 1(C) illustrate systems having very similar characteristics so they will be considered together. They are the most widely used systems at the present time and have given satisfactory service in many installations. With the advent of high-altitude, high-speed flight, however, certain undesirable characteristics appear which indicate that either modifications must be made on these systems or entirely new systems with better characteristics must be developed.

For example, at high altitude the

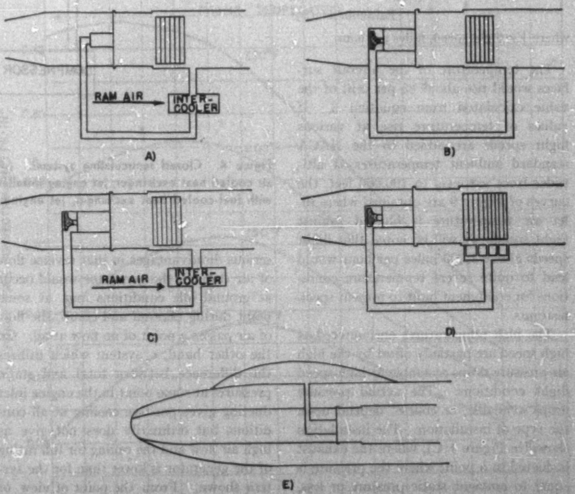


Figure 3. Bleed air cooling systems. (A) Compressor bleed, intercooler; (B) Compressor bleed, expansion turbine; (C) Compressor bleed, intercooler and expansion turbine; (D) Compressor bleed, expansion turbine with selective pressure take-off; (E) Cabin bleed

quantity of air that can be put through these machines is reduced considerably. This is illustrated on Figure 8 where the per cent of sea level air weight flow is plotted against altitude. For a fan-cooled machine, the volume handled would be approximately constant regardless of altitude and the weight flow of air then would be directly proportional to the density. If, as required by most specifications, the pressure drop across a blast-cooled machine is assumed to remain constant, the weight flow of air will vary approximately as the square root of the density. The curves shown are based on these assumptions.

It should be kept in mind, however, that most present generator specifications[5–7] were written for propellor-driven aircraft where the minimum available pressure drop would tend to remain constant with altitude. This is not the case for jet engine aircraft which tend to fly near maximum speed constantly regardless of altitude. For this situation, the available pressure drop and weight flow tend to fall off proportional to the density as for the fan-cooled machine.

Also, flight at high speeds causes a rise in cooling air inlet temperature, which in extreme cases might eliminate any possibility of using conventional blast systems. The rise in air temperature due to flight speed may be calculated from

$$\text{Temperature rise} = 1.8\left(\frac{V}{100}\right)^2 \text{ degrees Fahrenheit} \qquad (3)$$

where V = flight speed, miles per hour

The temperature of the aircraft surfaces would rise about 85 per cent of the value calculated from equation 3. If values of temperature rise at various flight speeds are added to the NACA standard ambient temperatures at altitudes from sea level to 100,000 feet, the curves of Figure 9 are obtained where inlet air temperature is plotted against flight speed. It will be noted that flight speeds above 1,000 miles per hour would lead to quite severe temperature conditions for equipment built to present specifications.

The high temperatures encountered at high speed are partially offset by the high air pressure drops available at high-speed flight conditions. The actual pressure drops available, of course, depend upon the type of installation. The installation shown in Figure 1(C), where the exhaust is ducted to a point where the pressure is equal to ambient static pressure or less, gives the greatest pressure drop across the machine at high-speed flight conditions. However, this type of installation has serious disadvantages in that reverse flow of air through the generator would occur at ground idle conditions and, at some point during take-off and climb, the flow of air passes a point of no flow at all. On the other hand, a system which utilizes the difference between total and static pressures at some point in the engine inlet ducting gives positive cooling at all conditions but ordinarily does not give as high air flow and the ceiling for full rating of the generator is lower than for the system shown. From the point of view of the generator, a combination of these two systems would be best for high-altitude, high-speed aircraft. It is possible to calculate the difference between total pressure generated during flight and ambient static pressure. These pressures were computed from the charts of reference 8, and are shown in Figure 10 as a function of flight speed and altitude. The actual pressure drop across the generator of Figure 1(C) can be higher or lower than the values shown depending upon the point at which the air exit is installed on the airplane.

If these large pressure drops are applied undiminished across a generator, the weight of cooling air and the drag on the airplane are quite high. An approximation of the airplane drag may be calcu-

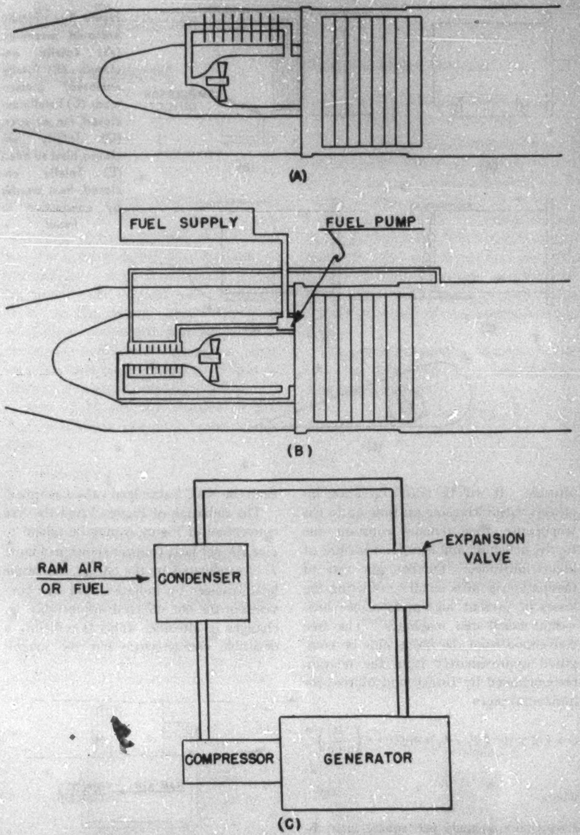


Figure 4. Closed recirculating systems. (A) Recirculating liquid or high-pressure gas with air cooled heat exchanger, jet engine installation; (B) Recirculating liquid or high-pressure gas with fuel-cooled heat exchanger, jet engine installation; (C) Mechanical refrigeration system

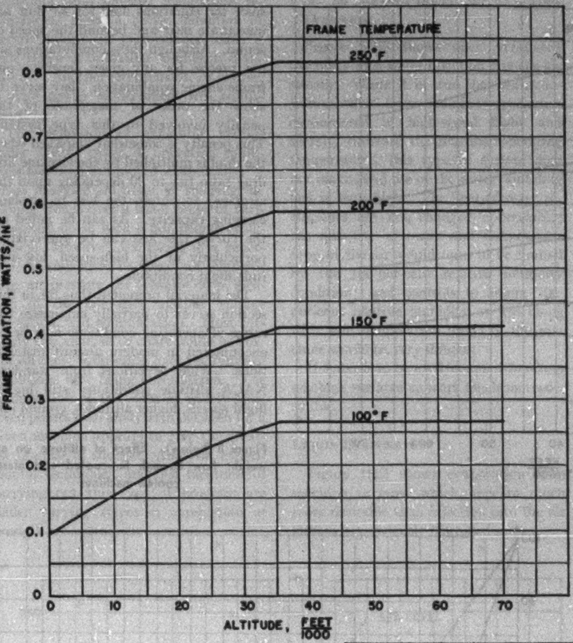


Figure 5 (above). Effect of altitude and frame temperature on radiation heat transfer. 100-per-cent emissivity; NACA standard temperature

lated as follows:

Assuming that the velocity of the air leaving the machine is negligible compared to the velocity of the plane

$$\text{Drag} = 1.467\frac{W}{G}V \qquad (4)$$

where

W = air weight flow, pounds per second
G = acceleration of gravity, feet per second^2
V = velocity of plane, miles per hour

The air weight flow, $W \sim \sqrt{P\Delta P}$ approximately (5)

where

P = average air density
ΔP = air pressure drop

For a 400-ampere 30-volt d-c machine present specifications allow 145 cubic feet per minute (0.185 pound per second) at sea level with 6 inches of water pressure drop across the machine. For such a machine, the weight flow at any other condition becomes, from equation 5

$$W = W_0\left[\frac{P(P_1-P_2)}{P_0(P_1-P_2)P_0}\right] \qquad (6)$$

but

$$P \sim \frac{P_{av}}{T_{av}} \sim \frac{P_1+P_2}{T_1+T_2} \qquad (7)$$

where

P_{av} = average pressure
P_1 = inlet pressure = ambient plus ram rise
P_2 = exit pressure = ambient
T_{av} = average absolute temperature, degrees Rankine
T_1 = absolute inlet temperature = ambient plus ram rise
T_2 = absolute exit temperature = inlet plus machine rise

T_2 may be defined from the losses of the machine and the air flow by

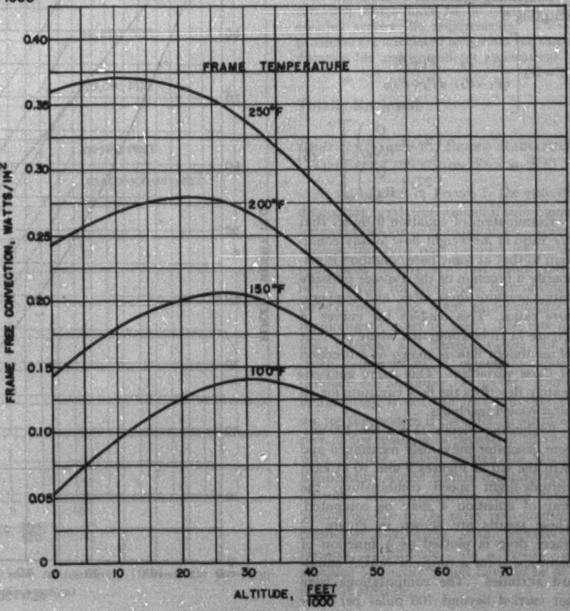
$$Q = WC_p(T_2 - T_1)$$

$$T_2 = \frac{Q}{WC_p} + T_1 \qquad (8)$$

where

Q = machine losses, British thermal units per second
C_p = specific heat of air, British thermal units per pound—degrees Rankine

Figure 6 (below). Effect of altitude and frame temperature on frame free convection. 8-inch diameter frame; NACA standard temperatures



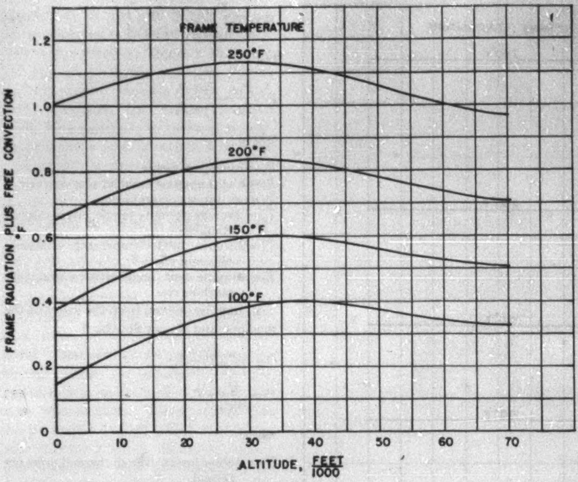


Figure 7 (above). Effect of altitude and frame temperature on the sum of frame radiation and free convection. 100 per cent emissivity; 8-inch diameter frame, NACA standard temperatures

Combining equations 7 and 8 and substituting 6

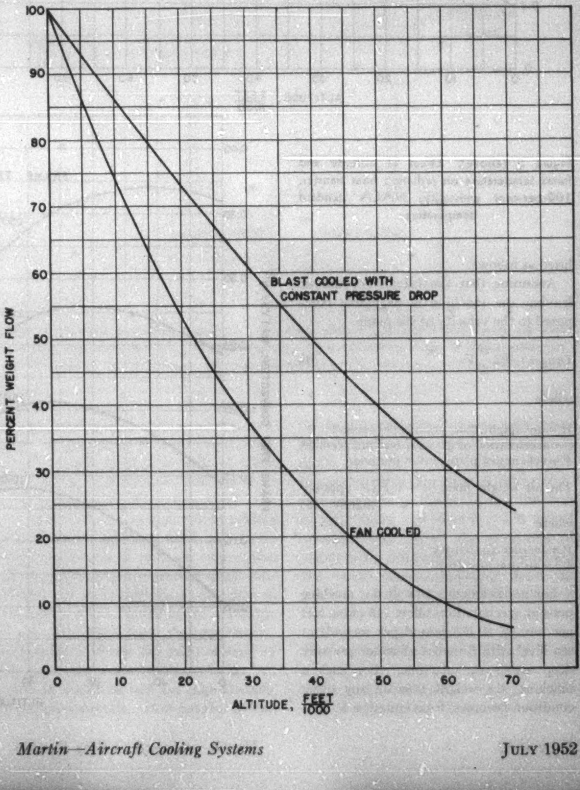
$$W = W_0 \left[\frac{P_1-P_2}{(P_1-P_2)}\ \frac{P_1+P_2}{{}_0(P_1+P_2)_0} \times \frac{\left(2T_1+\dfrac{Q}{WC_p}\right)_0}{\left(2T_1+\dfrac{Q}{WC_p}\right)} \right]^{1/2} \qquad (9)$$

Examination of equation 9 shows that the ratio of air weight flow at any condition to that at some base condition is primarily a function of the average pressure ratio, pressure drop ratio, and temperature ratio. This enables the curves of Figure 11 to be computed where per cent of air weight flow is shown as a function of these variables. This curve approximately describes the flow variation for all blast-cooled machines.

Evaluating Figure 11 for the 400-ampere generator previously mentioned and with the aid of Figures 9 and 10 at each altitude-flight speed combination, the drag of equation 4 may be computed. These results are shown in Figure 12 where drag is plotted as a function of flight speed for a range of NACA standard altitudes. The computations were not carried beyond 700 miles per hour since the equations used for air flow are inaccurate near and beyond the speed of sound. Although these computations are not precise for any given generator-airframe-engine combination, they serve to show the order of magnitude of the penalty involved in this type system. This penalty is somewhat amplified when the drag is multiplied by the airplane lift-drag ratio (up to 20 depending upon the airplane) to obtain the loss in airplane carrying capacity. As can be noted on the curve, this loss can be appreciable, particularly at the high-speed, low-altitude flight conditions.

The material presented thus far in this section serves to partially emphasize the range of operating conditions now being encountered in modern aircraft installations. Wide deviations from standard NACA altitude conditions, still higher flight speeds, higher altitudes, ground idle

Figure 8 (below). Effect of altitude on air weight flow through fan-cooled and blast-cooled machines



conditions, and unusual installation requirements make a better definition of electric equipment performance under a wide range of operating conditions mandatory. Another requirement, of course, is more nearly environment free equipment.

The definition of generator ratings under a wide range of operating conditions has been attempted analytically.[1–9] These analyses are useful in showing the cooling trends over a small range of variables and do so with a fair degree of accuracy. However, the rating of a generator depends upon its speed, air temperature, air pressure, air pressure drop, location of the hot spot, output voltage limitation, and commutation. Most certainly any analysis or chart would not define all these sufficiently over the range of conditions now being encountered. The combination of test data at widely scattered points with analysis in between for a given machine appears to give the greatest promise of success. Analysis, experimental techniques, and test facilities for carrying out this type of program are under varying stages of completion at present.

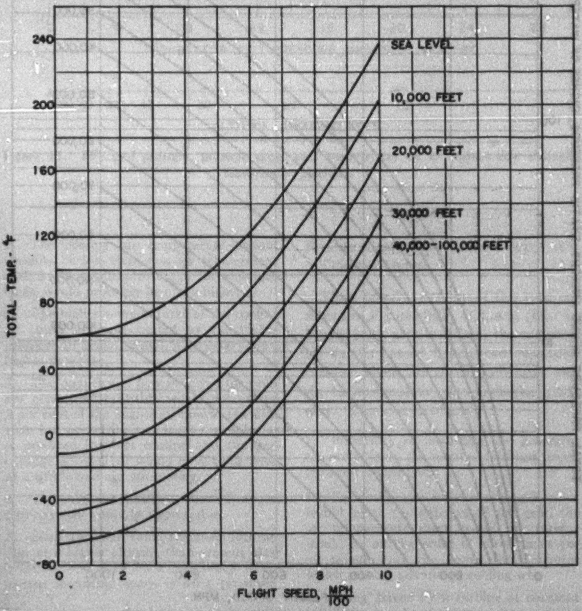


Figure 9. Effect of flight speed and altitude on total temperature. 100-per-cent recovery; NACA standard temperatures

Fan- or Blast-Cooled with Water Injection

Figure 1(D) shows open ventilated machines with water injection into the air stream. While it is true that water injection would lower the air temperatures encountered in high-speed flight and thereby decrease the machine operating temperatures, this type of system does not make good use of the water, and drag loss is very high. The heat-transfer coefficients are not changed appreciably; the quantity of water and temperature drop of the air would tend to be limited by the 100-per-cent relative humidity condition; and controls to assure 100-per-cent relative humidity or less along with safe temperature limits of the machine would be very difficult.

If water is to be used, evaporation from machine surfaces appears much more desirable.

Liquid Evaporation

Figure 1(E) shows evaporation from machine surfaces which appears much more desirable than injection into the air stream as previously discussed since:

1. The heat-transfer coefficients for boiling water are 20 to 200 times that for air.

2. The high heats of vaporization tend to keep the weight of coolant required at a reasonably low value (3.4 pounds per hour per kilowatt of losses). For example, the weight of water required to cool a 30-volt 400-ampere machine of 65 per cent efficiency is only 22 pounds per hour. For flights of short duration, the total weight of coolant required would be much less than the lifting capacity taken away from the plane by blast cooling.

3. For applications to short duration or expendable flights, such as guided missile, no control is required beyond a simple orifice set to meet full load cooling requirements. The weight of additional coolant used in most cases would be less than the weight of a more accurate control.

4. For high-altitude applications, the temperature levels in the machine would drop due to the lower boiling temperatures at low pressures.

5. In some installations the water could be kept from freezing by storing it in a separate container within the oil reservoir, or else low freezing point evaporants may be used.

6. Since the machine may be insulated against high temperature surroundings, such a system appears to be more nearly environment free and as simple as any system yet brought to the attention of the author.

7. Offsetting the apparent advantages are design problems such as commutation, brush life, moisture-proof insulation, and corrosion which have to be overcome in meeting service life requirements. These problems are alleviated somewhat for short duration life and expendable applications.

Totally Enclosed

Heat Transfer by Frame Radiation and Free Convection, Figure 2(A)

As was shown in Figure 7, the rate of dissipation is relatively small if safe frame temperatures are to be maintained. Nonetheless, this type may be used for small equipment having low loss rates, low-power rating (or short-cycle intermittent duty ratings), and where explosion-proof requirements must be met. The rate of heat transfer, and thereby rating, for a given temperature is essentially independent of altitude. However, frame temperature is not the best criterion to use for this system if the internal air pressure of the machine is permitted to vary with altitude. Even though the frame temperature would remain constant with altitude, the temperature of the rotor would rise due to the rotor heat-transfer coefficient decreasing with decreasing pressure at altitude. As a result the machine ordinarily would have to be derated at altitude.

By maintaining sea level pressures within the machine at all altitudes, as shown in Figure 2(B) the rating could be maintained constant. If pressures greater than sea level ambient were maintained, slightly greater ratings would be feasible as compared to present sea level ratings for this type of system.

FAN OR BLAST AIR OVER THE FRAME, FIGURES 2(C) AND (D)

As was found for the case of frame radiation and convection, the frame convection and radiation for these systems are also well defined analytically for any given configuration. Such analysis will not be made here due to the wide range of possible variations. At low altitudes, the allowable ratings for this type of totally enclosed machine are much higher than for the machine discussed in Figure 2(A) due to the much greater heat-transfer rate obtainable with forced convection. As the altitude is increased, however, the forced convection falls off and the rating approaches that obtainable by radiation alone.

The blast-cooled version would give the best cooling due to the much higher pressure drops available and lower losses resulting from leaving off the power absorbing fan. However, airplane drag and weight of air plumbing would in most cases favor the self-cooled version.

CONDUCTION OF HEAT TO FRAME

Figure 2(E) illustrates a motor making metal-to-metal contact with the skin of the wing. Such installations can provide considerable extra cooling for small totally enclosed machines. At present, however, this is not a dependable source of heat dissipation for electric equipment manufacturers since a given machine is just as likely to be installed on a mounting pad having a high temperature as one at a low temperature. Full utilization of this mode of heat transfer will require careful joint study by airframe and electrical manufacturers to determine whether the heat flow path can be kept short enough, small enough, and light enough to be feasible. As pointed out by Eckert,[10] the most suitable materials for conducting the heat would be aluminum or magnesium. The density to conductivity ratios of these materials are much lower than any of the other common metals, resulting in a much lower weight for a given heat-transfer rate.

Another possibility along these lines is the flush skin heat exchanger analyzed for electronic equipment by Robinson and Zimmerman.[2] This system comprises the heat generating element, a liquid transfer medium, a pump, and a flush skin heat exchanger. The system is equally applicable to rotating machinery, particularly for motors, inverters, and auxiliary power units installed in remote parts of the airplane.

It should be kept in mind, however, that even this mode of transferring heat is not free of altitude effects. After being conducted to the airframe, the heat must still be convected to the air where the convection heat-transfer coefficient drops off at altitude. Further, in the supersonic flight speed range, the temperature of the airplane skin may prohibit high heat-transfer rates.

Bleed Systems

ENGINE COMPRESSOR

Many possible variations exist utilizing air from the compressor for driving and cooling the generator. First, however, it is well to keep two things in mind:

1. The temperature at the exit of the compressor is almost always much too hot to cool electric equipment with present temperature limitations, unless the air is cooled in some way.
2. As pointed out in reference 11, the thrust and fuel consumption are affected appreciably by air bleed. For the engine tested, it was found that 10-per-cent bleed air resulted in 77-per-cent lower thrust and 117-per-cent higher fuel consumption and was almost linear with per cent of bleed air at lower values. Thus, any discussion of any bleed air cooling system would have to take this penalty into consideration.

ENGINE COMPRESSOR BLEED WITH INTERCOOLER

Figure 3(A) illustrates one of the obvious systems. Some of the advantages and disadvantages of the system appear to be:

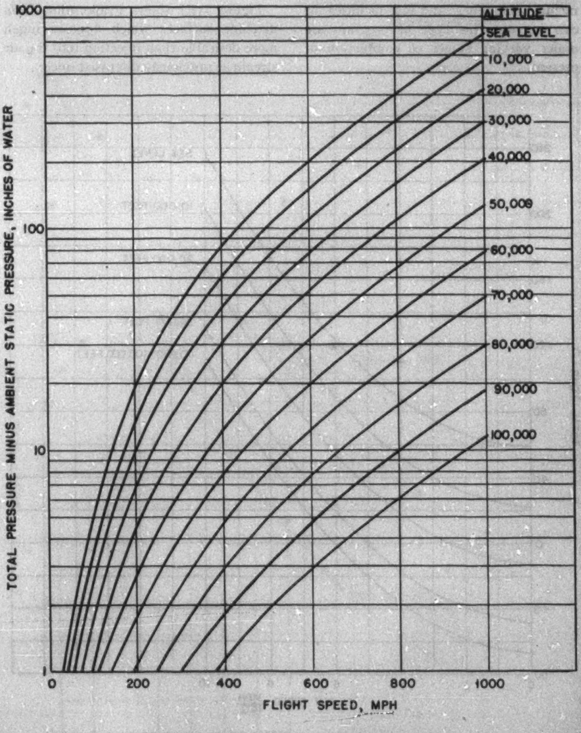


Figure 10. Effect of flight speed and altitude on ram air pressure. 100-per-cent recovery, NACA standard temperatures and pressures

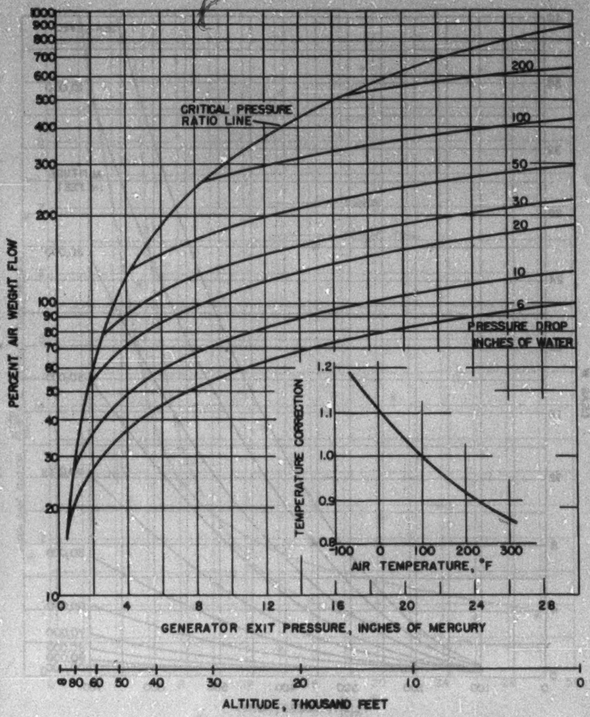


Figure 11. Effect of pressure, pressure drop, and temperatures on air weight flow through blast-cooled machines

1. The cooling air temperature cannot possibly be less than ram air temperature and, for an optimum design on a weight basis, would probably be much higher.
2. The higher temperatures of the cooling air could be compensated by the higher pressure air available and, if necessary, by a higher flow rate.
3. The quantity of flow involved for present generators probably would not exceed 1 per cent of the engine air flow at low altitude but even this is as large a penalty as is involved in blast cooling and the percentage of bleed air would have to increase at altitude to maintain cooling.
4. The weight of plumbing, heat exchanger, and controls would be appreciable.
5. Such a system could maintain full ratings at a higher altitude than present blast systems, and may be a possibility where engine mounting space is at a premium.

The penalties involved appear to be quite severe, however, and larger, better-cooled, more efficient blast-cooled generators appear to have better long-range possibilities.

Engine Compressor Bleed with Air Turbine Drive

Three possible systems of this type are shown in Figures 3(B), (C), and (D). In Figure 3(B), air taken from the compressor exit is passed through an expansion turbine which drives the generator and the turbine exit air, then cools the generator. This system offers:

1. Possibility of mounting the generator remotely from the engine, relieving overcrowded engine pads.
2. Since the efficiency of the generator would be on the order of 65 per cent, the air temperature drop across the turbine would be on the order of three times the air temperature rise across the generator which is a big advantage cooling-wise.
3. Being driven by a turbine at constant speed, the weight of the generator could be appreciably reduced although a d-c machine would have to be driven through a gear box due to extremely high turbine speeds. This could become a severe problem but is being done in some instances at the present time.

For the normal pressure ratios and flight speeds of jet engine aircraft, however, the compressor outlet temperature may still be too high for effective cooling after leaving the expansion turbine. This may be overcome by the system shown in Figure 3(C) where a heat exchanger is placed ahead of the expansion turbine. This heat exchanger may utilize either fuel on its way to the burner or ram air, whichever appears most feasible from analysis of a given installation.

The over-all efficiency of air turbine installations utilizing air taken from the compressor exit is very low since the turbine must be designed for the pressures available at the engine idle condition. Then, at normal operating conditions, the control must throttle the air to much lower pressures resulting in poor use of the available energy in the air. This condition may be eliminated to a large extent by using a selective compressor pressure take-off control, as indicated in Figure 3(D), rather than a throttling control. That is, bleed air off at the compressor exit during idle, then switch progressively to earlier stages of compression as the engine comes up to speed. If the airplane flies fast enough, it may even be possible to switch to the engine inlet diffuser and utilize the ram air to drive the air turbine. For an installation of this type, the drive efficiency of the air turbine system would approach that of conventional gear systems. It further offers the advantages of constant speed drive and adequate cooling at no extra cost. Detailed analysis would show whether and when an intercooler were necessary for the system shown in Figure 3(D).

Bleed Air from Pressurized Cabin, Figure 3(E)

Very little data are available at present on the range of air pressures, temperatures, and quantities of air that are available for this type of cooling. Indications are that the temperatures are too high, the pressures too low, and the quantities of air too low for cooling present generator designs. However, this system should not be overlooked on this basis since a slightly heavier, more efficient, higher heat-transfer area machine may do an adequate job in this type of installation at a lower over-all penalty on the airplane than the present blast-cooling systems. Further, as flight speeds continue to rise, the need arises for greater capacity cabin cooling systems which should supply greater quantities of air for accessory cooling systems.

Closed Recirculating Systems

Several possibilities exist here such as:

1. Recirculating liquid or gas with an air-cooled heat exchanger, as shown in Figure 4(A).
2. Recirculating liquid or gas with a fuel-cooled heat exchanger, as shown in Figure 4(B).
3. Mechanical refrigeration system, as shown in Figure 4(C).
4. In addition, several other possibilities exist which could be considered such as utilizing a flush skin heat exchanger, de-icing system heat exchanger, or even a cabin heater in some cases.

All these versions are environment bound to some degree although the mechanical refrigeration system could be designed for almost any given environment.

RECIRCULATING GAS OR LIQUID SYSTEMS WITH AIR HEAT EXCHANGER

These systems are environment bound in the same fashion as present blast-cooled systems. This is due to the decrease in available pressure drop and air density at high altitudes which would require a larger, heavier heat exchanger. Also, at high-flight speeds, the cooling air temperature at the heat exchanger can become so high as to preclude any possibility of cooling. Still further, these systems introduce an extra barrier to the flow of heat as compared to conventional blast-cooled machines. The only advantages which appear possible for this type are: a tendency to eliminate hot spots in the machine; and possibility of reducing the altitude brush wear problem by maintaining a controlled atmosphere around the brushes. In general, it appears that blast cooling is inherently better than this configuration.

FUEL HEAT-EXCHANGER SYSTEM

This system, shown in Figure 4(B), would tend to require a larger heat exchanger for the sea level engine idle and high altitude operating conditions since the engine fuel flow rate is reduced considerably at these conditions. However, the high heat-transfer coefficients obtainable with liquids would keep the heat exchanger weight very small even for these conditions. The extremely high fuel flow rates of the propulsion systems in modern high performance aircraft make this type of system appear very attractive. Some possible advantages are:

1. Possibly more reliable brush performance since this type of system yields itself to cooled brush boxes and totally enclosed, controlled atmosphere brush compartment.
2. Positive cooling at all altitudes and flight speeds whenever the engine is operating.
3. Lower penalty from the airplane than for blast cooling; the slightly extra weight would be much less serious than the air pickup drag loss now being encountered in blast cooling.
4. Elimination of air piping and engine compressor entrance disturbances.
5. Makes possible a cooling system that can be made an integral part of the power plant.
6. Susceptibility to contaminants and fire hazard are eliminated since this would be a totally enclosed machine.
7. Rejects heat at moderately high temperatures and is therefore not as sensitive to heat sink temperatures as blast cooled machines.
8. The high heat-transfer coefficients obtainable with liquids would serve to reduce the commutator temperatures encountered in high rotative speed operation.
9. There is also the possibility of using a recirculating liquid which evaporates in the generator or other rotating machine and condenses in the fuel heat exchanger. This, if feasible, would serve to reduce the required weight flow of the coolant and the required heat-transfer areas.

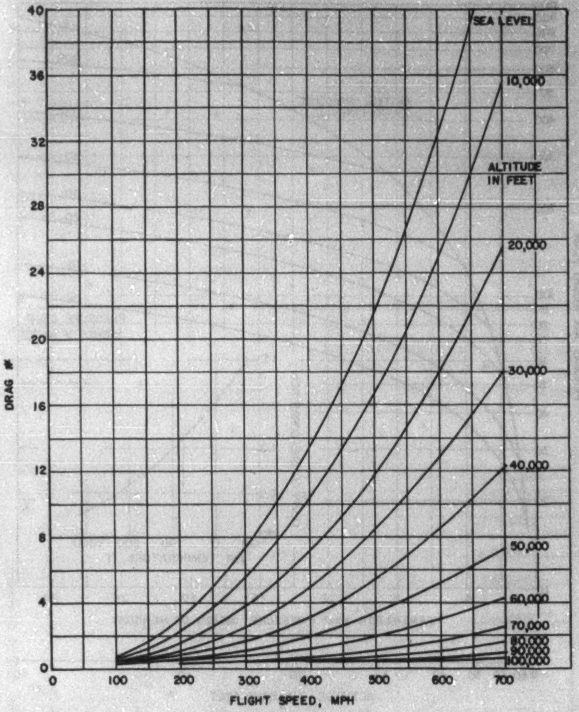


Figure 12. Effect of altitude and flight speed on pickup drag of a 400-ampere 40-volt military specification generator

Despite these apparent advantages, the fuel-cooled system may not be applicable unless sufficient quantities of a fuel are available at moderate or low temperatures. To determine this the following data from engine and airframe manufacturers are required before an installation can be made.

1. Minimum weight flow of fuel versus altitude at outlet of the fuel pump.
2. Maximum generator load versus altitude.
3. Maximum temperature and maximum allowable temperature of fuel at outlet of last pump versus altitude.
4. How much heat from other sources is being put into fuel after last pump.

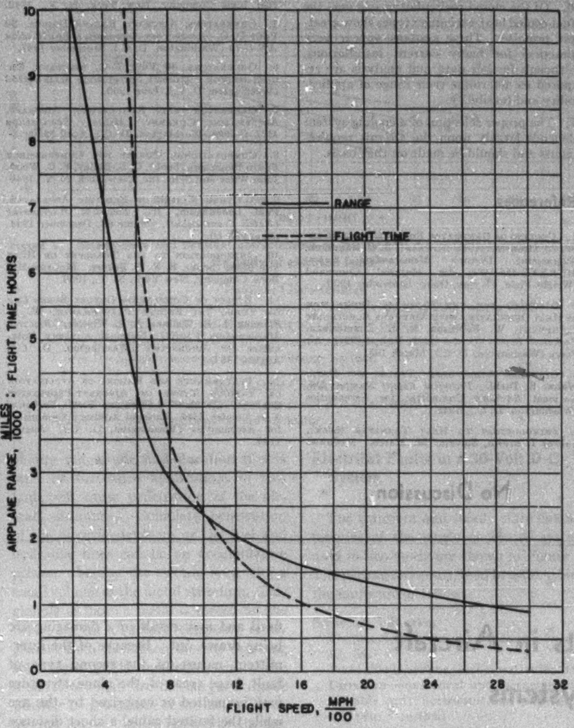


Figure 13. Effect of flight speed on maximum airplane range and flight time

5. If the fuel is now being used to the fullest extent for cooling, could the pressure, and thereby the boiling temperature, be increased enough to cool the generator?

Mechanical Refrigeration System

This system, shown in Figure 4(C) does not appear feasible for most applications, because:

1. The weight of the system would be quite high.
2. The additional complexity would add to maintenance and life requirement difficulties.
3. The space required probably would not be available on most engines.
4. Almost all applications where the temperature conditions were severe enough to justify this system also could be cooled by the evaporative system of Figure 1(E). Further, if the temperatures were that high, the design of refrigeration components would present severe problems.
5. However, if such a system were found to be necessary for cooling the cabin, the fuel or both, it could be made slightly larger to accommodate the generator.

Higher Temperature Limits

From the standpoint of cooling system weight, higher temperature limits are desirable regardless of the type of cooling system selected. The altitude and flight speed range of applicability of all systems may be increased with increased temperature limits. Machines with higher temperatures must have higher copper losses due to higher resistance.

The accomplishment of higher temperature limits requires:

1. Better brushes.
2. Better insulation.
3. Elimination of hot spots.
4. Greater strength on rotating parts.
5. Greater thermal stability, particularly for commutators.

Selection of a Cooling System

The primary considerations in the intelligent selection of a cooling system for an aircraft installation are so rigidly tied to required aircraft performance that they cannot be separated. To even begin, there are such things as:

1. Maximum operating altitude.
2. Maximum flight speed, and maximum temperature when greater than that due to ram.
3. Maximum operating time for the electric equipment.
4. Duty cycle for the electric equipment must be known.

From the material presented in reference 12, an envelope curve of maximum probable airplane range versus flight speed was drawn and is shown as Figure 13. From the same data, it is possible to determine an approximate flight time-speed curve which is also shown. The propulsion systems vary along the curve in the following manner:

Flight Speed, Miles per Hour	Power Plant
100– 500	Reciprocating or turbine propeller
500–1,200	Turbojet
1,200–1,600	Turbo-ram jet
1,400–2,000	Ram jet
2,000 and up	Rocket

Not shown, but also of interest, is an extremely rapid increase in rate of fuel consumption as the flight speed is increased. Thus, as the required aircraft flight speed goes up, the time of flight goes down to the point where expendable evaporative cooling looks more attractive.

Further, since the fuel flow rates increase with flight speed, fuel-cooled systems also become more desirable. And at the extremely high flight speeds and short duration flight the thermal capacity of the machine may be great enough that no cooling at all is required.

It should be kept in mind, however, that propulsion systems and aircraft for high-altitude and high-speed flight have undergone tremendous changes, and changes of the same order of magnitude can be expected. For example, the development of nuclear power plants for aircraft or a satellite rocket could change Figure 13 to almost infinite range and flight time at almost any flight speed. Needless to say, such developments would certainly impose new conditions for electric equipment and their cooling.

Conclusions

1. Present blast-cooled, fan-cooled, and totally enclosed machines give satisfactory service in the great majority of present aircraft. However, at high altitude and/or high speed, other systems will be needed and will extract a lower over-all penalty from the airplane than present systems.

2. Expendable evaporative cooling appears to offer the greatest promise of essentially environment free operation. It also shows promise in replacing blast cooling in applications such as short duration flight.

3. Totally enclosed machines should be pressurized wherever possible for high-altitude operation. They should also be installed so as to take advantage of conduction heat transfer to the aircraft frame.

4. Air turbine driven generators using the turbine exhaust for cooling show advantages that should be investigated further, particularly with selective compressor pressure take-off at moderate altitudes.

5. Cabin bleed cooling appears to be limited to low output machines at the present time.

6. Of the closed recirculating systems, the fuel-cooled heat-exchanger types show greatest promise. These systems appear very practical for many current installations, although further data and analysis are required to determine their range of applicability and feasibility.

7. The proper selection of a cooling system depends largely upon the aircraft requirements and should be made on that basis.

References

1. Cooling of Generator Equipment for High-Speed, High-Altitude Aircraft, **E. O. Naumann.** *Engineering' Division Memorandum Report MCREXE-656-1898*, Air Material Command (Wright Field, Dayton, Ohio), September 1950.

2. Methods for the Ultimate Dissipation of Heat Originating with Airborne Electronic Equipment, **W. Robinson, R. H. Zimmerman.** *Technical Report Number 6439*, United States Air Force (Washington, D. C.), March 1951.

3. Standard Atmosphere—Tables and Data, **Walter S. Diehl.** *Technical Report Number 218*, National Advisory Committee for Aeronautics (Washington, D. C.), 1948.

4. Introduction to Heat Transfer (book), **Aubrey I. Brown, Salvator M. Marco.** McGraw-Hill Book Company, New York, N. Y., 1942.

5. Generators, Aircraft, Engine Driven, 30 Volt D. C. *Army-Navy Aeronautical Specification AN-G-1a* (Washington, D. C.), September 1947.

6. Generators, 30 Volt D-C, Aircraft, Engine Driven. *Military Specification MIL-G-6162* (Washington, D. C.), June 1950.

7. Generator and Regulators, Aircraft, Alternating Current. *Military Specification MIL-G-6099* (Washington, D. C.), April 1950.

8. Computational Curves for Compressible Fluid Problems (book), **C. L. Dailey, F. C. Wood.** John Wiley and Sons, Inc., New York, N. Y., 1949.

9. Altitude Rating of Electric Apparatus, **Paul Lebenbaum, Jr.** *Electrical Engineering* (*AIEE Transactions*), volume 63, December 1944, pages 955–60.

10. Introduction to the Transfer of Heat and Mass (book), **E. R. G. Eckert.** McGraw-Hill Book Company, New York, N. Y., 1950.

11. Effect of Compressor Outlet Bleed Off on Turbo Jet Engine Performance, **W. A. Fleming, L. E. Wallner, J. R. Wintler.** *Research Memorandum E50E17*, National Advisory Committee for Aeronautics (Washington, D. C.), August 1950.

12. Performance and Ranges of Application of Various Types of Aircraft-Propulsion Systems, **Cleveland Laboratory Staff.** *Technical Note Number 1349*, National Advisory Committee for Aeronautics (Washington, D. C.), August 1947.

No Discussion

Intermittent Faults in Aircraft Electric Systems

L. R. LARSON MEMBER AIEE

M. TRBOVICH ASSOCIATE MEMBER AIEE

P. E. TOOMIRE ASSOCIATE MEMBER AIEE

IN most electric systems in military aircraft, a grounded installation is used, that is, the conducting parts of the plane structure form the electrical return path to the generator. The cables may be bundled and secured to the plane structure for support with the cable insulation preventing metal-to-metal contact. In an installation of this kind, contact between an energized cable and the plane structure with current flowing through this junction constitutes an electrical fault.

Electrical faults may be considered as being either of two types depending upon the nature of the contact at the point of the fault. The first type of fault is characterized by continuous metal-to-metal contact between the electric conductor and the plane structure. The contact resistance may be high or low in magnitude depending upon the number of conductor strands in contact, the contact pressure, and the contamination of the contact surfaces. This type of fault usually results in damage to the electric system. To minimize damage from this type of fault, the fault-sensing device operating time should be correlated with the smoking time characteristic of the conductor being protected or with the current-time characteristic of equipment being protected.

A second type of fault in an aircraft electric system is characterized by separation of a faulted cable and the aircraft structure after contact has been made. Such separation may be intermittent or fixed and may result in a damaging arc being drawn out. Because of the intermittent nature of the second type of fault, large areas of the plane structure may be melted or vaporized by the arc while the faulted cable, a short distance from the arc, is not excessively hot.[1–4]

An intermittent fault may be described by the current magnitude and fault pulse duration. However, the fault current magnitude and duration depend upon the number of wire strands in contact with the plane structure, the contact pressure of each strand, the contact area, the variation of arc length with time, the surface contamination, and other parameters. To describe faults which exist in aircraft, it is necessary to know the exact manner in which the fault is established. Such information is not available and it is assumed that electrical faults in aircraft can be caused by parting a conductor in any number of ways. To determine the approximate minimum fault current magnitude and duration, many light-contact line-to-ground faults were established in simulated 30-volt d-c, 120-volt d-c, and 120/208-volt, 400-cycle a-c systems.

The First Fault Pulse

The first fault pulse was selected for study since it is desirable to detect a fault

---

Paper 52-137, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE South West District Meeting, St. Louis, Mo., April 15–17, 1952. Manuscript submitted August 3, 1951; made available for printing February 27, 1952.

L. R. Larson, M. Trbovich, and P. E. Toomire are all with the Naval Research Laboratory, Washington, D. C.

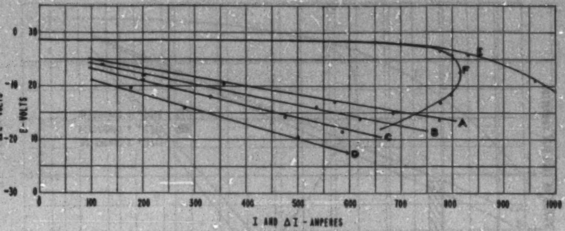


Figure 1. Steady-state and transient characteristics of the 30-volt d-c electric system used in the study of electrical faults

Transient Characteristics
Curve A—System composed of one generator and one battery, 200-ampere load
Curve B—System composed of one generator and one battery, no load
Curve C—System composed of one generator, 200-ampere load
Curve D—System composed of one generator, no load

Steady-State Characteristics
Curve E—System composed of one generator and one battery
Curve F—System composed of one generator

of any magnitude and duration if this fault or successive applications of this fault will cause penetration of the aircraft structure. Complete penetration of the aircraft structure or the fuel and hydraulic lines can be an accumulative process whereby the electric arc melts a small volume of the metal structure. This globule of molten metal is forced by the cable, by the structure vibration, or by gravity from the crater, thereby exposing solid metal to the succeeding arc. Such a series of small arc discharges is made possible by the nature of the insulation on aircraft cables. When ignited by the arc, this insulation burns and leaves a brittle residue. The residue then is separated from the conductor by vibration and thus more bare wire is exposed to continue the intermittent fault.

The data presented concern, for the most part, the first fault pulse of intermittent type faults. The magnitude and duration of the first fault pulse is approximately the same as the magnitude and duration of the succeeding pulses. The time between the first and second fault pulse is approximately the same as the time between any other two consecutive fault pulses in the intermittent fault.

Electrical Faults in a 30-Volt D-C System

The transient and steady-state characteristics of the single generator system used in this study are shown in Figure 1. The parameters considered in establishing the faults are as follows:

Structure Vibration
Frequencies, cycles per second—0; 15; 30; 50
Total displacement, inches—0.0125; 0.05
Direction—horizontal and perpendicular to cable run; horizontal and parallel to cable run; vertical

Batteries
Connected; disconnected

Faulted Cable
Size, number—2; 8; 16
Length, feet—5; 25

Structure Thickness, inches—0.05; 0.091

Distance from Clamp to Fault, inches—2; 8; 12

All of the established faults did not result in arcs. Of the 2,342 faults established, approximately 23 per cent resulted in welded faults and the remaining 77 per cent resulted in intermittent faults; see Table I.

The current magnitudes of the faults on each size conductor varied considerably. For example, the average fault current per pulse on size number 16 conductor varied from approximately 40 amperes to over 700 amperes. The average fault currents per pulse on sizes number 8 and number 2 conductors extended from 50 amperes to approximately 1,000 amperes.

Additional fault current magnitude and duration data for these faults are given in Figures 2 and 3.

Data from 622 faults established on conductors 5 feet long show a slightly greater number of higher current faults as the structure-vibration frequency increased from 15 to 50 cycles per second. In addition, structure-vibration frequency in this range has no marked effect on the duration of the first fault pulse.

Data from 208 faults established on cable size number 8 show a small change in the distribution of fault currents for a change in vibration displacement from 0.0125 inch to 0.05 inch. No effect on pulse duration was noted for this change in displacement.

Data from 619 faults established on conductors 5 feet long show that the direction of vibration has relatively little effect upon the current magnitude of the first fault pulse and causes no marked change in fault pulse duration.

Intermittent Faults in a 120-Volt D-C System

The transient and steady-state characteristics of the system used in this study are shown in Figure 4. The parameters considered in establishing these faults are as follows:

Batteries
Connected; disconnected

Structure Vibration
Frequencies, cycles per second—0; 30; 50
Total Displacement, inches—0.0125; 0.05
Direction—horizontal and perpendicular to cable run; horizontal and parallel to cable run; vertical

Faulted Cable
Size, number—2; 8; 16
Length, feet—10

Load Prior to Fault
No load; 100 amperes

Table I. Number of Intermittent Faults and Welds Which Resulted from Establishing 2,342 Faults in a 30.0-Volt D-C Aircraft Electric System

Plate Thickness, Inches	Battery On		Battery Off		AN Conductor Size Number 16		AN Conductor Size Number 8		AN Conductor Size Number 2				
	Number of Arcs	Number of Welds	Number of Arcs	Number of Welds	Number of Arcs	Number of Welds	Number of Arcs	Number of Welds	Number of Arcs	Number of Welds	Total Arcs	Total Welds	Total Faults
0.05	581	146	556	172	303	179	412	73	422	66	1137	318	1455
0.091	505	166	155	61	185	112	241	51	234	64	660	227	887
Totals	1086	312	711	233	488	291	653	124	656	130	1797	545	2342

Of the 180 faults established, 9 per cent resulted in welded faults and the remaining 91 per cent were of the intermittent type.

As in the 30-volt d-c study, there is considerable variation in fault current for each cable size. For example, the average current per fault pulse on size number 16 cables ranged from 210 amperes to approximately 500 amperes; the average current on size number 8 cables ranged from approximately 275 amperes to over 800 amperes; and fault current on size number 2 cable ranged from approximately 315 amperes to over 800 amperes; see Figure 5.

An increase in structure-vibration frequency from zero to 50 cycles per second results in a slight increase in the number of higher current faults.

Data from 52 intermittent faults established on cable size number 8 show that there is a tendency toward a greater number of higher current faults as the displacement is increased from zero to 0.0125 inch and 0.05 inch. The direction of vibration has no marked effect on the fault-current magnitude.

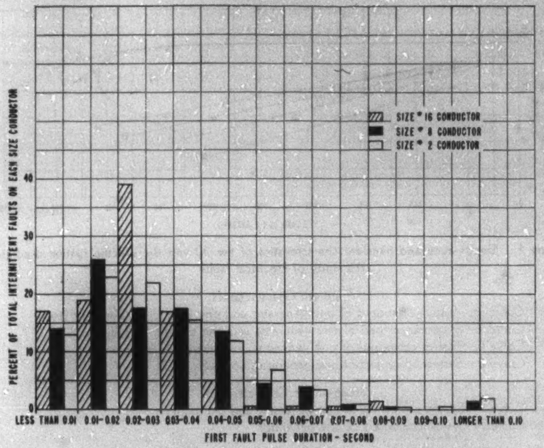


Figure 3. Duration of the first fault pulse of 614 faults established in a 30-volt d-c system between cables 5 feet long and an aluminum alloy plate 0.05-inch thick

To define further the intermittent faults established on a simulated aircraft electric system, it is necessary to know the duration of the first fault pulse; see Figure 6. Cable size has no pronounced effect on first fault pulse duration although there is a slight increase in duration as the conductor cross-sectional area is increased. Also structure-vibration frequency has little effect on the duration of the first fault pulse. In addition, structure-vibration displacement has little effect on the duration of the first pulse of fault current. Direction of structure vibration has only a minor effect on the duration of the first fault pulse.

Of the 180 faults established, 41 of these faults caused complete penetration of the aluminum alloy plate within 1 second.

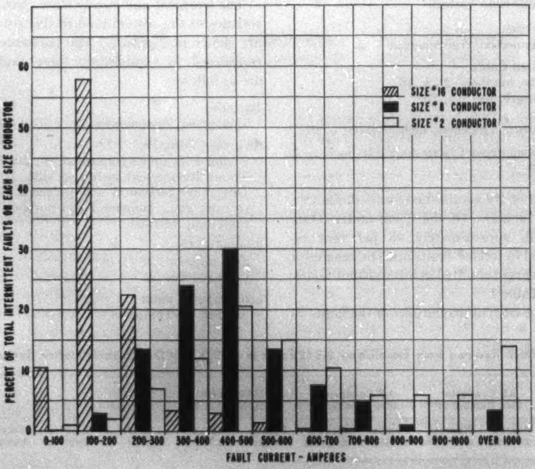


Figure 2. Current magnitude of the first fault pulse of 621 faults established in a 30-volt d-c system between cables 5 feet long and an aluminum alloy plate 0.05-inch thick

Intermittent Faults in a 120/208-Volt 400-Cycle System

The electric power source used for this study is a 40-kva aircraft generator driven by a 50-horsepower motor. Only single-phase faults were established. The parameters considered when faults were established are as follows:

Structure Vibration
  Frequencies, cycles per second—0; 30; 50
  Total Displacement, inches—0.0125; 0.05
  Direction—horizontal and perpendicular to cable run; horizontal and parallel to cable run; vertical

Faulted Cable
  Size, number—8; 12; 16
  Length, feet—10

Load Prior to Fault
  Type—no load; balanced; unbalanced
  Power factor—1.00; 0.75 (lag)

Of the 225 faults established, approximately 5 per cent developed into welded faults; however, no welds resulted for faults established on a nonvibrating platform. As would be expected, test results show that although maximum fault currents occur on the cable with the smaller resistance, a wide range of fault currents were measured on each of the three sizes of cables used; see Figure 7.

The first fault pulse duration of 98 per cent of the total intermittent faults is

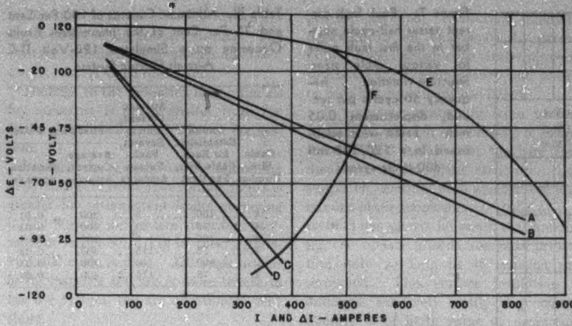


Figure 4. Transient and steady-state characteristics of the electric system used in this study

Transient Characteristics

Curve A—System composed of one generator and four batteries, 100-ampere load
Curve B—System composed of one generator and four batteries, no load
Curve C—System composed of one generator, 100-ampere load
Curve D—System composed of one generator, no load

Steady-State Characteristics

Curve E—System composed of one generator and four batteries
Curve F—System composed of one generator

less than 0.01 second, and none of the faults established penetrated an aluminum alloy plate 0.091-inch thick within 1 second.

Analysis of Fault Currents

It was found that for a given fault the average magnitude of the current pulses following the first is approximately the same as for the first pulse. Using the time between pulses, the fault pulse magnitude, and duration, the average fault current can be calculated. These calculated values for faults on a 30-volt d-c system are shown in Table II. Similarly, in the 120-volt d-c system, the average current of the first current pulse of all of the faults struck on size number 16 cable is 200 amperes or greater. The first pulse duration of all of the faults struck on size number 16 cable is 0.01 second or greater, and the time between the first and second faults is 0.3 second or less. These values for number 16 cable give the minimum magnitude of the faults occurring on the simulated 120-volt d-c system used in this study. The average fault current, considered over the duration of several pulses, is calculated to be 5 amperes. The average current of the first fault current pulse of 75 per cent of the faults struck on size number 16 cable is 300 amperes or greater and the first pulse duration of 75 per cent of the faults struck on size number 16 cable is 0.01 second or greater. The time between the first and second fault pulse of 75 per cent of the faults is 0.05 second or less. From these values, considered over the duration of several pulses, the average fault current is calculated to be 50 amperes. Similar data were obtained for cable sizes number 8 and number 2, and the results are shown in Table III. Comparison of the values in Tables II and III shows that the 120-volt system does not give appreciably higher minimum fault currents than the 30-volt system.

Due to the extremely wide variation in magnitude of the first half-cycle of fault current in the a-c system, it was omitted from this analysis. The average duration of the first fault pulse was $2^1/_2$ cycles

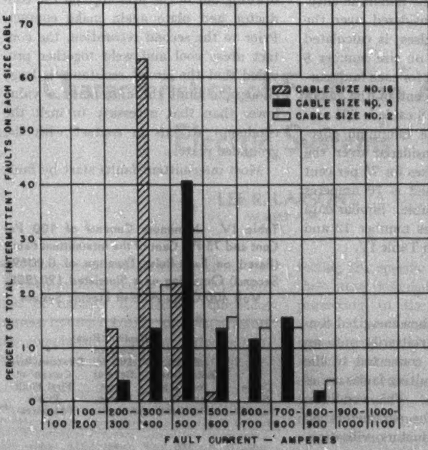


Figure 5. Current magnitude of the first fault pulse of 163 faults established in a 120-volt d-c system between cables 10 feet long and an aluminum alloy plate 0.091-inch thick

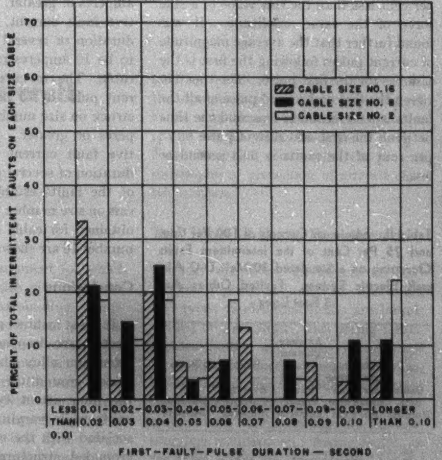


Figure 6. Duration of the first fault pulse of 85 faults established in a 120-volt d-c system between cables 10 feet long and an aluminum alloy plate 0.091-inch thick

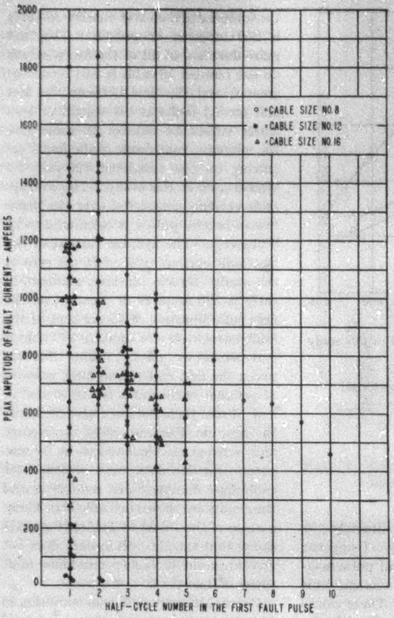


Figure 7. Peak fault current versus half-cycle number in the first fault pulse for various cable sizes. Structure-vibration frequency 50 cycles per second, displacement 0.05 inch. Faults were established in a 120/208-volt 400-cycle system

(0.0063 second). For single-phase faults on the particular generator used, the rms value of fault current is approximately 10 per cent less than the rms value of a sine wave of the same amplitude. It was found further that the average magnitude of current pulses following the first is the same as for the first pulse. The time between the first and second pulse of all the faults is 0.08 second or less, and the time between the first and second pulse of 75 per cent of the faults is 0.04 second or less.

Table II. Minimum Currents of 100 Per Cent and 75 Per Cent of the Intermittent Faults Occurring on a Simulated 30-Volt D-C Aircraft Electric System. Faulted Cables Are 5 Feet Long

Cable Size, Number	Faults Established for Each Cable Size, Per Cent	Average Current Due to Several Fault Pulses, Amperes	First Fault Pulse: Average Current, Amperes	First Fault Pulse: Duration, Seconds
16	100	5	50	0.005
16	75	25	100	0.01
8	100	5	50	0.005
8	75	75	300	0.01
2	100	5	50	0.005
2	75	100	400	0.01

With this interpretation, the rms current for each fault current pulse of all the faults struck on size number 8 cable is 150 amperes or greater. The minimum effective fault current, considered over the duration of several pulses, is calculated to be 10 amperes rms on size number 8 cable. The rms current for each fault current pulse of 75 per cent of the faults struck on size number 8 cable is 505 amperes or greater. The minimum effective fault current considered over the duration of several pulses for 75 per cent of the faults established is 70 amperes rms on size number 8 cable. Similar data obtained for cable sizes number 12 and number 16 are shown in Table IV.

Conclusions

In most instances, when energized conductors are dropped arbitrarily onto an aluminum alloy plate connected to the system ground, the resulting faults are of the intermittent type. This condition exists while varying the parameters associated with the conductors, vibration, grounded structure, system load, and system voltage. Further, the minimum fault currents encountered are not changed when the system voltage is increased but depend upon the resistance of contact, as determined by the area of the contact, and the arc resistance. It was found further that the fault-current magnitude and duration of the first fault pulse is not markedly affected by vibration frequency, displacement, and direction although fault duration increases when the faults are established on a vibrating platform.

Table III. Minimum Currents of 100 Per Cent and 75 Per Cent of the Intermittent Faults Occurring on a Simulated 120-Volt D-C Aircraft Electric System

Cable Size, Number	Faults Established for Each Cable Size, Per Cent	Average Current Due to Several Fault Pulses, Amperes	First Fault Pulse: Average Current, Amperes	First Fault Pulse: Duration, Seconds
16	100	5	200	0.01
16	75	50	300	0.01
8	100	5	200	0.01
8	75	115	400	0.02
2	100	10	300	0.01
2	75	115	400	0.02

There are two conditions under which welds occur. First, if there is sufficient contact pressure to prevent separation of the conductors of opposite polarity, and further, if there is sufficient circuit resistance to limit the current to a value lower than that necessary to melt the contact strands, the electrical fault will be of the welded type. Second, the conductor strands are heated during the drawing out of the first arc and the conductor and plate again make contact. Prior to the second separation, the contact areas cool and weld together provided that the circuit resistance is large enough to limit the current to a value lower than that necessary to melt the conductor strands in contact with the grounded plate.

Most intermittent faults start by fault

Table IV. Minimum Currents of 100 Per Cent and 75 Per Cent of the Intermittent Faults (Based on Fault-Pulse Duration of 0.00625 Second) Occurring on a Simulated 120/208-Volt 400-Cycle Aircraft Electric System

Cable Size, Number	Faults Established for Each Cable Size, Per Cent	Effective (Rms) Current Due to Several Fault Pulses, Amperes	Effective (Rms) Current of First Fault Pulse, Amperes
8	100	10	150
8	75	70	505
12	100	15	200
12	75	60	460
16	100	5	70
16	75	55	390

current flow in excess of the current carrying capacity of the contact strands. The contact strands become molten and part, causing an arc to be drawn out.

Damage to the grounded aluminum alloy structure is much greater from the faults established in the 120-volt d-c system.

Further, the damage to the aluminum alloy structure from light contact faults is greater for intermittent faults occurring in the 30-volt d-c system than for intermittent faults occurring in the 120/208-volt 400-cycle system. Although in this latter instance the arc is more severe in the a-c system, the arc duration is very short.

A fault-clearing device designed to clear all of the intermittent-type faults, and based upon the average fault current over several fault pulses, should clear a fault current of approximately 10 amperes. With present techniques in protection, however, this seems impractical. A fault-sensing device designed to clear a fault based upon the magnitude and duration of a single fault pulse appears to be more practical. A relay method of fault detecting and clearing, however, necessarily limits fault protection to a limited region of the aircraft. This is true because of the space and weight requirements for the numerous items required to protect each cable. In addition, it may be desirable in an emergency to permit faults on electric cables to vital equipment to exist until the emergency has passed or the vital equipment has failed.

If complete protection from ground faults is required, it is necessary to insure the electrical isolation of the wire conductors from the plane structure. One suggested method is to anodize the aluminum plane structure. A study of this method has shown, however, that anodizing the surface affords adequate protection only so long as it remains unscratched. This surface insulation is damaged relatively easily and consequently is of little value as protection against electrical faults.

Another means of preventing the occurrence of some electrical faults consists of the installation of an ungrounded system monitored by a ground detector. This method, from a protection standpoint, would have merit if the possibility of simultaneous multiple grounds on cables of different polarity and line-to-line faults are disregarded.

Another means of gaining limited protection from faults is to provide cables able to withstand the conditions leading to a sustained intermittent fault. First, this requires a cable insulation which is fire resistant when heat is applied from the conductor side of the insulation. Second, a cable insulation is needed which is secured to the conductor in such a manner so as to expose a minimum of bare wire when the cable is parted. Third, an abrasion-resistant cable insulation is required.

It is recognized that isolating electrical faults can be accomplished in limited portions of the aircraft electric system. Further, with the present relaying techniques and aircraft equipment requirements, a protection system should be committed to clear on the first fault pulse of an intermittent fault in order to minimize structural damage. For the large majority of wire footage, however, fault preventive measures seem to be the most practical means to minimize fault damage in aircraft electric systems.

References

1. Characteristics of Faults on Aircraft Electric Systems, **Lee R. Larson.** *AIEE Transactions*, volume 71, part II, July 1952, pages 165–69.

2. Electric Contacts, **Ragnar Holm.** H. Gebers, Stockholm, Sweden, 1946.

3. Heat Effects in Anode Spots of High Current Arcs, **T. B. Jones, W. B. Kouwenhoven, M. Skolnik.** *Welding Journal* (New York, N. Y.), volume 28, October 1949, pages 4615–55.

4. A-C and D-C Short-Circuit Tests on Aircraft Cable, **J. C. Cunningham, W. M. Davidson.** *AIEE Transactions*, volume 63, December 1944, pages 961–69.

No Discussion

Characteristics of Faults on Aircraft Electric Systems

LEE R. LARSON
MEMBER AIEE

THE protection of electric systems has been applied mainly to the prevention of damage to electric equipment caused by overheating or overstressing the components. Normal industrial installations result in systems that, in protecting the specific equipment for which they were designed, also give adequate protection to the surrounding equipment. Unfortunately, it has been assumed, without serious reservation, that similar protective systems for aircraft equipment could perform the same function. In reality, however, the environmental conditions in aircraft operation are much more severe—that is, the structure of the aircraft is much more vulnerable to arc damage—requiring that more extensive and efficient protection be used in the future. It would be rather pointless to install a system that assured the operation of specific equipment if, at the same time, it allowed the structure of the aircraft to be destroyed. Such a contingency can occur under the present philosophy of protection of aircraft electric systems.

Two Types of Faults Encountered

Under the severe environmental conditions imposed by aircraft operation, two types of electrical faults are found: welded or solid faults, and arcing faults. Welded faults, which are characterized by high, sustained values of current, limited mainly by the circuit resistance, cause extensive damage to the electric system with minor damage to the surrounding installations. Arcing faults, on the other hand, are characterized by the intermittent flow of current with magnitude limited by the arc and circuit resistance and the transient characteristics

---

Paper 52-138, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE South West District Meeting, St. Louis, Mo., April 15–17, 1952. Manuscript submitted October 26, 1949; made available for printing February 27, 1952.

Lee R. Larson is with the Naval Research Laboratory, Washington, D. C.

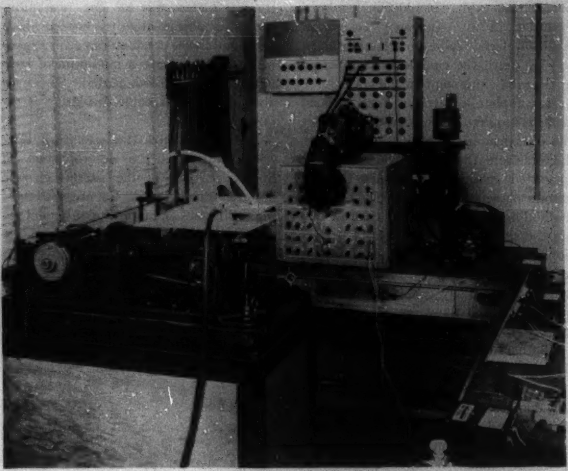


Figure 1. Laboratory installation for the study of arcing faults showing the method of mounting the aluminum ground plane and the cable used to establish the fault

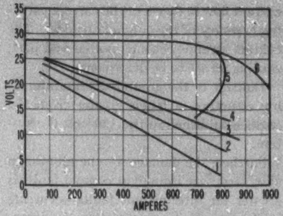


Figure 3. Steady-state and transient characteristics of a 30-volt 400-ampere aircraft generator

Transient characteristics

1. One generator, no load
2. One generator, 200-ampere load
3. One generator and battery, no load
4. One generator and battery, 200-ampere load

Steady-state characteristics

5. One generator
6. One generator and battery

of the power source; these faults may cause severe structural damage with little effect on the electric system.

WELDED FAULTS

The philosophy that faults are self-clearing has long since been discarded. With high-capacity generating systems various types of protection have been used. With overvoltage protection, differential protection, reversed polarity protection, automatic equalizer disconnects, and other types, the electric system is well attended. These types of equipment normally will remove a faulted section without serious damage to the electric equipment. Since these methods of protection have been well described in the literature and are applicable to welded faults, no further mention will be made of welded faults.

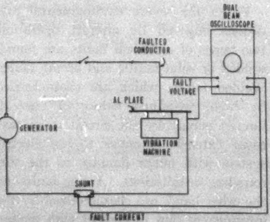


Figure 2. Schematic diagram of the circuit used in the study of arcing faults

ARCING FAULTS

Insufficient attention has been given to faults giving rise to continuous or intermittent arcing. Laboratory tests indicate average current levels of a few amperes on circuits ranging from number 18 to 1/0 cable. Obviously, the problem of detecting such small average current differentials on circuits capable of carrying several hundred amperes will be exceedingly difficult. To obtain data that would yield useful information on the characteristics of arcing faults, the following experimental arrangement was used. The physical arrangement of the ground plane and the faulted conductor on a vibration table is shown in Figure 1. The associated circuit arrangement is shown schematically in Figure 2. In laboratory tests a vibration table frequency of 30 cycles per second and a displacement of 0.050 inch was used as an approximation of aircraft vibration. Brief surveys of the effect of varying frequency and displacement were made prior to selecting these values. Vibration was applied vertically and horizontally with minor variations in fault characteristics. The data to be presented were accumulated using horizontal vibration.

System Characteristics

The steady-state and transient characteristics of the 30-volt aircraft generator[1] are given in Figure 3. These characteristics were taken at the point of application of the faults. It is apparent from these characteristics that the power source does influence the fault current though the variation is obscured by the variable nature of the arc resistance during the fault.

To simplify the test program the oscillograms were taken without load or battery connected to the bus.

The generator supplying power for faults above 30 volts was excited by a controlled exciter having considerable time delay. The approximate steady-state and transient characteristics of this

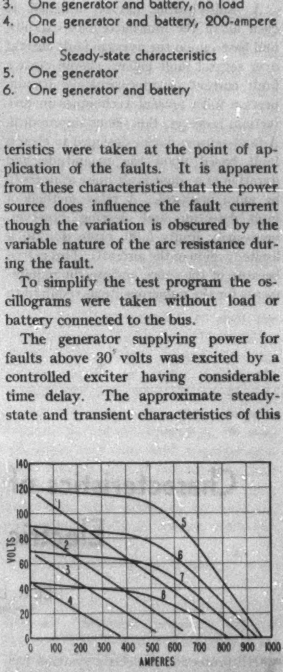


Figure 4. Steady-state and transient characteristics of variable voltage source 200-ampere 12- to 120-volt d-c generator

Transient characteristics

1. 120 volts, no load
2. 90 volts, no load
3. 70 volts, no load
4. 45 volts, no load

Steady-state characteristics

5. 120 volts
6. 90 volts
7. 70 volts
8. 45 volts

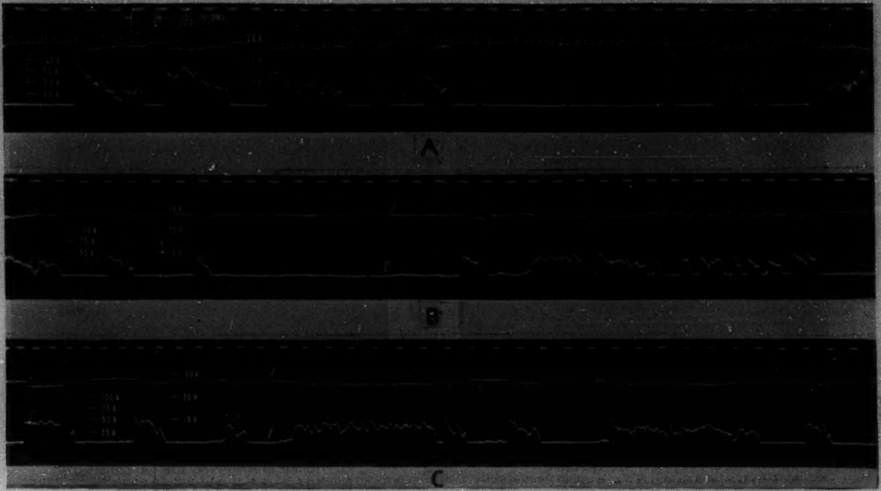


Figure 5. Oscillograms showing the characteristics of typical arcing faults using one 30-volt 400-ampere aircraft generator. The ground plane was 0.050-inch aluminum

A. 15 feet of Army-Navy size 10 cable
B, C. 10 feet of Army-Navy size 16 cable

generator at the point of fault are given in Figure 4. Characteristics are given for 45, 70, 90, and 120 volts.

Arcing Fault Characteristics

For 30-volt systems, the nature of the arcing fault often results in intermittent arcing. As can be seen in Figure 5, the peak value of the current is limited by the circuit resistance. The normal arc current is considerably below this value and is determined by the arc resistance and the transient characteristics of the supply system. Visually, one has the impression that the arcing is continuous. The oscillogram shows conclusively that there is no current conduction for a large portion of the time. The nature of the damage done

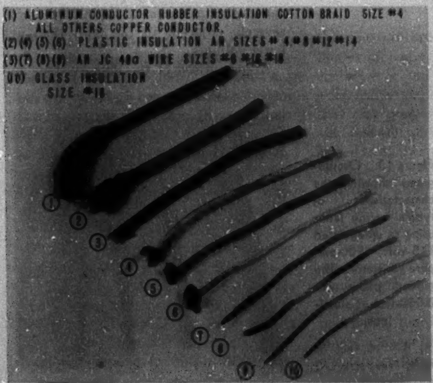

(1) ALUMINUM CONDUCTOR RUBBER INSULATION COTTON BRAID SIZE #4
ALL OTHERS COPPER CONDUCTOR.
(2)(4)(5)(6) PLASTIC INSULATION AN SIZES #4 #8 #12 #14
(3)(7)(8)(9) AN JC 48a WIRE SIZES #6 #16 #18
(10) GLASS INSULATION SIZE #18
①
②
③
④
⑤
⑥
⑦
⑧
⑨
⑩

Figure 6. Typical damage to cables by arcing faults at 30 volts

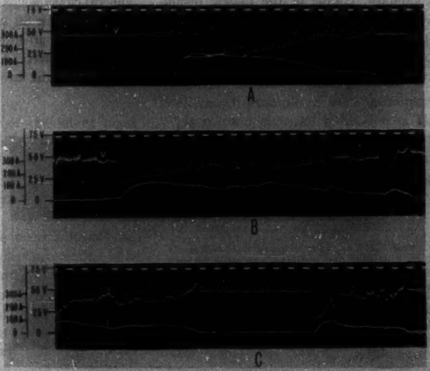


Figure 7. Oscillograms showing the characteristics of typical arcing faults using an unregulated 45-volt 200-ampere generator. The ground plane was 0.050-inch aluminum

A. 10 feet of Army-Navy size 1/0 cable
B. 10 feet of Army-Navy size 8 cable
C. 10 feet of Army-Navy size 16 cable

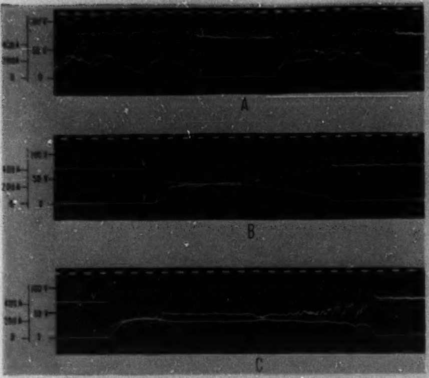


Figure 8. Oscillograms showing the characteristics of typical arcing faults using an unregulated 70-volt 200-ampere generator. The ground plane was 0.050-inch aluminum

A. 10 feet of Army-Navy size 1/0 cable
B. 10 feet of Army-Navy size 8 cable
C. 10 feet of Army-Navy size 12 cable

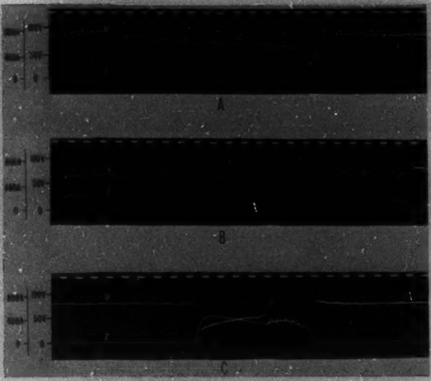


Figure 9. Oscillograms showing the characteristics of typical arcing faults using an unregulated 90-volt 200-ampere generator. The ground plane was 0.050-inch aluminum

A. 10 feet of Army-Navy size 1/0 cable
B. 10 feet of Army-Navy size 8 cable
C. 10 feet of Army-Navy size 12 cable

to cable by such faults can be seen in Figure 6. From Figure 6, it is apparent that the type of insulation does influence the arc. The damage to the insulation is due to heat conduction along the conductor from the arc. In burning, rubber insulation forms an enlarged, rigid, charred material that tends to prevent further contact between the ground plane and the conductor. Some of the plastic insulations exhibit a self-healing characteristic in that the plastic softens and may seal off the conductor during a period of nonconduction. The normal Army-Navy cable has the undesirable characteristic of forming a thin, brittle residue after burning. In the presence of vibration, this material is broken back from the end of the conductor tending to enhance the arcing fault.

On going to higher system voltages, two marked phenomena are noticed. As would be expected, the arc is much more violent causing considerably more damage. Between 30 and 40 volts, the arcs change from an intermittent nature to a continuous arc. Due to the extreme violence of the arc for 120-volt equipment, the ground plane may be burned back until the fault is cleared. However, since the conductor is exposed to continued vibration, wind load, or other action, it would re-establish the arc on contact with any grounded metal.

Typical oscillograms of faults at increased voltages are shown in Figures 7, 8, 9, and 10. It will be noted that the flow of current into the fault is more stable for the higher voltages, indicating increased average power consumption in the arc. Damage to faulted conductors is indicated in Figure 11. The high-power dissipation causes the insulation to be burned more severely adjacent to the fault than at lower voltages.

A comparison of the damage done to

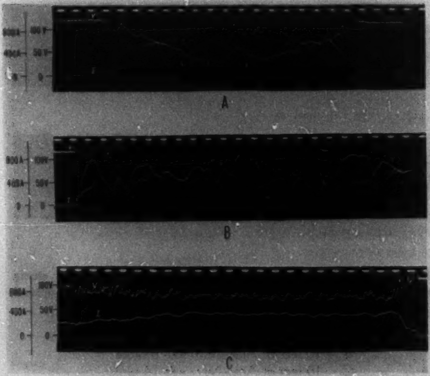


Figure 10. Oscillograms showing the characteristics of typical arcing faults using an unregulated 115-volt 200-ampere generator. The ground plane was 0.050-inch aluminum

A, B. 10 feet of Army-Navy size 1/0 cable
C. 10 feet of Army-Navy size 4 cable

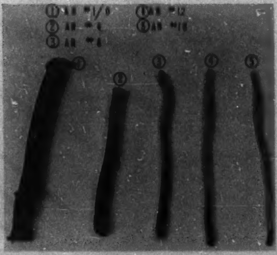


Figure 11. Typical damage to cables by arcing faults at 115 volts

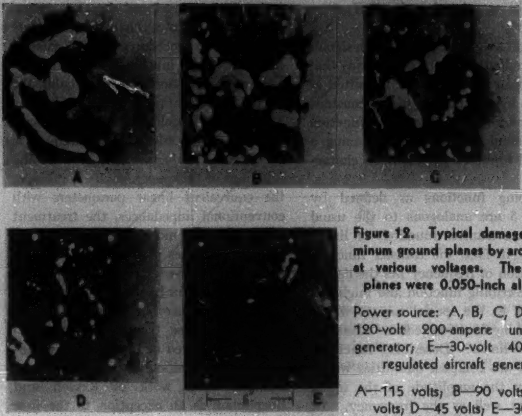


Figure 12. Typical damage to aluminum ground planes by arcing faults at various voltages. The ground planes were 0.050-inch aluminum

Power source: A, B, C, D—12- to 120-volt 200-ampere unregulated generator; E—30-volt 400-ampere regulated aircraft generator

A—115 volts; B—90 volts; C—70 volts; D—45 volts; E—30 volts

aluminum plates simulating the aircraft structure is given in Figure 12. It will be noted that slots are burned in the plates in many cases. This seems to be associated with the phenomenon of charred insulation breaking off under vibration as mentioned earlier. These plates are typical of the damage done by the various system voltages. Each plate was used for a number of tests at the particular voltage.

Conclusions

The premise that any fault endangering the safety of the aircraft must be cleared leads to the following conclusions:

1. Arcing faults may not overload the individual circuit.
2. Structural damage and fire hazard due to arcing faults may be extensive.
3. System protection based on welded fault conditions is not adequate.
4. Differential protection, as now applied to generator circuits, is not adapted easily to the entire distribution system due to the increased weight and vulnerability.
5. Consideration must be given to the use of an ungrounded system to decrease the probability of encountering an arcing fault.

Reference

1. D-C Power Systems for Aircraft, Part V, R. H. Kaufmann, H. J. Finison. *General Electric Review* (Schenectady, N. Y.), volume 49, May 1946, pages 46–54.

No Discussion

Sinusoidal Analysis of Feedback-Control Systems Containing Nonlinear Elements

E. CALVIN JOHNSON
ASSOCIATE MEMBER AIEE

TECHNIQUES commonly used at the present time in the analysis and synthesis of feedback-control systems are largely those of constant-linear system theory. In the usual form, these methods stem from a system characterization in terms of ordinary linear differential equations having real constant coefficients. Fortunately, for many purposes this characterization is satisfactory, and in these cases the properties and behavior of the systems can be studied analytically by application of the large body of available linear theory. Frequently, however, practical control systems contain elements, either unavoidably or by design, which are not linear in the engineering sense of being describable by linear differential equations with constant coefficients.

Reliance on linear procedures in the first place prevents the fullest investigation and utilization of the latent possibilities of truly nonlinear control. The contactor or "on-off" servomechanism is a familiar example of a nonlinear system that may possess striking advantages in size and simplicity over a system designed to do the same job on a linear basis. However, even in systems intended to be essentially linear, incidental nonlinearities such as those caused by saturation, backlash, deadspace, and variable damping inevitably are present if the systems are examined closely enough. These parasitic effects often cause characteristically nonlinear behavior that cannot be understood thoroughly and evaluated by studying linear differential equations with constant coefficients, which wholly ignore the responsible nonlinearities.

Because of the mathematical difficulties involved, a completely general, all-inclusive theory of nonlinear systems appears to be virtually impossible at the present time. Nevertheless, much can be learned about nonlinear systems through more or less specialized techniques that yield approximate information in certain fairly restricted situations. The frequently used linearization procedure, the so-called "method of small motions," is a well-known example of such techniques. This method gives an indication of the behavior of a system for small amplitude variations about a fixed operating point. Such information can be quite useful and far from trivial. In fact, this is the basis

---

Paper 52-154, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted February 20, 1952; made available for printing April 10, 1952.

E. Calvin Johnson is with the Bendix Aviation Corporation, Detroit, Mich.

The material presented in this paper is based upon results obtained as part of a doctorate thesis research program conducted by the author at the Massachusetts Institute of Technology. The opportunity provided by the Dynamic Analysis and Control Laboratory for carrying out this work is deeply appreciated. The author wishes to acknowledge with sincere gratitude the many helpful suggestions offered during the course of the research by several members of the Institute staff, particularly Professor R. H. Frazier, who supervised the thesis.

on which most vacuum-tube circuits are designed.

The sinusoidal analysis which is the subject of this paper is another example. Its most important use is in determining whether or not there is danger of sustained oscillation or destructive instability in a system, and in showing how such behavior can be avoided or the effects diminished. The method also is helpful in providing an indication of simple transient behavior. But the method should be thought of as a special technique rather than as a general solution to the general nonlinear problem. It is fairly general in the sense that it is applicable to a large and important class of nonlinear effects, but it is not very useful, for example, in finding the response of a system to random excitation.

Describing Functions

The basis of the method is the assumption that the input or driving force imposed upon a nonlinear element is sinusoidal and that the resulting response of the element can be represented satisfactorily, insofar as the performance of the system as a whole is concerned, by its fundamental Fourier component. A so-called "describing function" is defined as the amplitude and phase angle of the fundamental component of the response relative to the assumed driving sinusoid. Thus, suppose that the output x_k of an element is given as a function of its input x_j by the relation

$$x_k = g_j(x_j) \tag{1}$$

Assume now that the input x_j is sinusoidal, say

$$x_j = ReX_j\epsilon^{i\omega t} \tag{2}$$

Elements for which this analysis is useful produce an output when sinusoidally excited which is periodic with a period equal to that of the input sinusoid. Consequently, when the input has the form of equation 2, the output x_k can be represented by the Fourier series

$$x_k = \frac{1}{2}\sum_{m=-\infty}^{\infty} X_{km}\epsilon^{im\omega t} \tag{3}$$

in which

$$X_{km} = \frac{1}{\pi}\int_{-\pi}^{\pi} g_j(x_j)\epsilon^{-im\omega t}d\omega t \tag{4}$$

The describing function N_j for the element g_j is defined by

$$N_j = X_{k1}/X_j \tag{5}$$

The component of fundamental frequency in x_k therefore is $ReX_{k1}\epsilon^{i\omega t} = ReN_jX_j\epsilon^{i\omega t}$.

Constant average components in signals can be handled simply by considering that the nonlinear characteristic is shifted or displaced by the amount of the constant component. The describing function then is found as shown by assuming a sinusoidal input to the modified characteristic.

Describing functions as defined by equation 5 are analogous to the usual steady-state transfer functions of linear elements; in fact, the transfer function as a function of frequency fits the definition of describing function and might be thought of as the special case to which the describing function reduces when the element is linear. Describing functions in general are complex quantities which depend on the amplitude of the input sinusoid as well as the frequency. In many cases of interest, however, they depend on amplitude alone in contrast to steady-state transfer functions which depend only on frequency. For the purpose of studying system behavior, describing functions for the nonlinear elements are combined with conventional transfer functions for the linear portions of the system. In this way, many of the familiar techniques of sinusoidal analysis can be extended to nonlinear systems.

The scheme of representing a nonlinear element by a describing function or its equivalent has been reported by several writers as a means for investigating both oscillators and control systems. In the latter category, the work of Kochenburger contains an extensive treatment of the method, particularly as applied to contactor servomechanisms.[1] Describing functions are developed for several contactors, and combination of these describing functions with the linear transfer functions of the remainder of the systems for use in Nyquist diagrams of conventional servo theory is discussed and illustrated fully.

Certain methods of a more mathematical nature, developed in the general field of nonlinear mechanics, are closely related to the procedures under discussion in that they involve essentially the same approximations. In particular, the "equivalent linear parameters" derived by Kryloff and Bogoliuboff from consideration of a second-order differential equation are directly analogous to the describing functions of this paper.[2,3] By combining the equivalent linear parameters with conventional impedances, the treatment was extended to more complex systems.

The techniques presented by Kochenburger and others are of considerable practical value in handling nonlinearities in feedback-control systems. For the most part, these techniques appear to have been arrived at intuitively, and independently, incidentally, by the various workers. The work of Kryloff and Bogoliuboff is an exception to the first part of this statement in that their equivalent linear parameters as applied to second-order systems are derived from a mathematical investigation. As a general method, however, the techniques seem not to have been developed in sufficient generality and analytical detail to permit adequate evaluation of a number of significant topics. The work described here is an attempt to fill partially this need by providing a more satisfactory mathematical foundation for the describing-function method. One of the important results is a means either for estimating and investigating the error of the analysis or for obtaining a more accurate solution for a general system containing n nonlinear elements.

Illustrative Application of the Method

Before undertaking a detailed mathematical discussion, practical use of describing-function analysis will be illustrated through application to a simplified aircraft roll-control system containing a kind of backlash or free play in the control-surface linkage. The nature of the system is indicated schematically in Figure 1. A gyroscope measures the difference between a desired roll angle ϕ_i

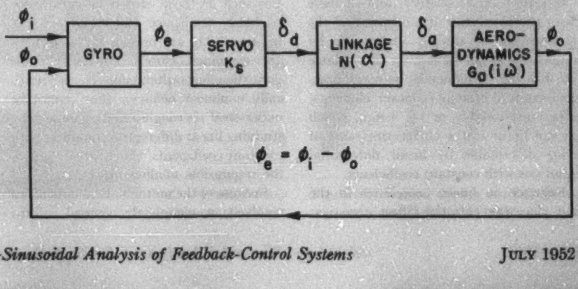


Figure 1. Simplified aircraft roll-control system

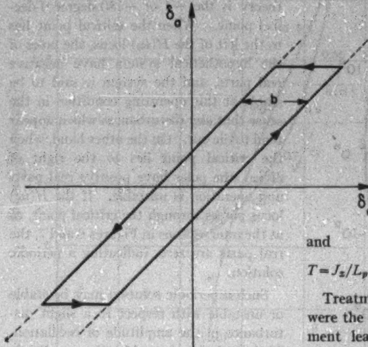


Figure 2 (left). Free-play or backlash characteristic

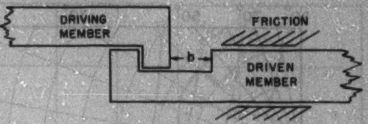


Figure 3 (right). Possible model for the free-play characteristic

and the actual roll angle ϕ_0. The difference or error ϕ_e is applied to a servomechanism which actuates the ailerons through a mechanical linkage. Deflection of the ailerons δ_a produces a moment about the roll axis, causing roll motion in approximate accordance with the equation

$$J_x\frac{d^2\phi_0}{dt^2}+L_p\frac{d\phi_0}{dt}=L_\delta\delta_a \tag{6}$$

in which J_x, L_p, and L_δ are aerodynamic coefficients.

The control-surface servo is taken to be ideal; that is, servo output δ_d is given by

$$\delta_d=K_s\phi_e \tag{7}$$

However, free play having the characteristic shown in Figure 2 is assumed to be present in the linkage between δ_d and δ_a. A possible model for such a characteristic is suggested in Figure 3. For sinusoidal δ_d, δ_a is periodic having the form sketched in Figure 4. The amplitude and phase of the fundamental component of δ_a may be computed in terms of a, the amplitude of δ_d, and b, the amount of free play. For convenience, let α be the nondimensional amplitude ratio defined by

$$\alpha=a/b \tag{8}$$

The describing function $N(\alpha)$, giving the amplitude and phase angle of the fundamental component of δ_a relative to δ_d, is thus determined and may be sketched in the complex plane as in Figure 5. The aerodynamic transfer function $G_a(i\omega)$, obtained by conventional theory is

$$G_a(i\omega)=\frac{\Phi_0(i\omega)}{\Delta_a(i\omega)}=\frac{K_a}{i\omega(i\omega T+1)} \tag{9}$$

in which

$$K_a=L_\delta/L_p \tag{10}$$

and

$$T=J_x/L_p \tag{11}$$

Treatment of $N(\alpha)$ just as though it were the transfer function of a linear element leads to the sinusoidal response characteristic of the over-all system

$$\frac{\Phi_0(i\omega)}{\Phi_i(i\omega)}=\frac{K_sN(\alpha)G_a(i\omega)}{1+K_sN(\alpha)G_a(i\omega)} \tag{12}$$

$$=\frac{N(\alpha)H(i\omega)}{1+N(\alpha)H(i\omega)} \tag{13}$$

with $K_sG_a(i\omega)=H(i\omega)$. This expression is exact whenever δ_d is sinusoidal; but even if δ_d is only approximately sinusoidal, equation 13 may be sufficiently accurate to give useful results. The assumption that nonlinear elements have sinusoidal driving functions is, as already stated, basic in this method of analysis.

Consideration of equation 13 and the nature of N and H shows that self-sustained oscillations may exist when

$$1+N(\alpha)H(i\omega)=0 \tag{14}$$

or when

$$H(i\omega)=-\frac{1}{N(\alpha)} \tag{15}$$

The functions $H(i\omega)$ and $-1/N(\alpha)$ are sketched in the complex plane in Figure 6 and in the "gain-phase" plane* in Figure 7 for a particular combination of system constants which produce (without the free play) a system having a nondimensional damping ratio of 0.2. The intersection observed in these figures corresponds to the solution of the condition equation 15. It thus represents a possible periodic solution to the equations describing the system.

As already mentioned, describing functions may depend on frequency as well as amplitude. In these cases, the relation analogous to equation 15 should be written

$$H(i\omega)=-\frac{1}{N(\alpha,i\omega)} \tag{16}$$

in which α again is a measure of the amplitude. The function $-1/N(\alpha,i\omega)$ in general must be plotted as a family of loci. One form that these loci might take is illustrated in Figure 8. Intersections representing solutions of equation 16 are those at which the frequencies of the $H(i\omega)$ and the $-1/N(\alpha,i\omega)$ loci are the same. Two such intersections, at ω_2 and ω_3, are shown in Figure 8. When the describing functions depend on both amplitude and frequency, location of significant intersections is more difficult. But, as the analysis presented in the Appendix indicates, the general method is just as valid as when the system can be separated into amplitude invariant and frequency invariant portions.

In general, single-valued nonlinear characteristics, such as those of saturation, deadspace, and the like, lead to purely real describing functions. The $-1/N$ loci then lie along the 0-degree or -180-degree lines. Frequency-dependent describing functions arise when time is involved in some way, such as through time derivatives of input or output.

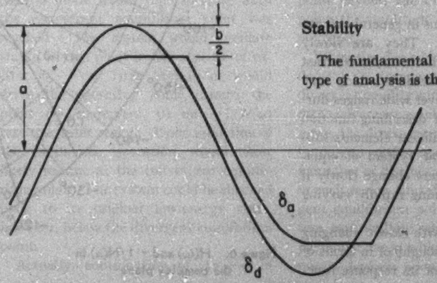


Figure 4. Wave forms of δ_d and δ_a

Stability

The fundamental idea underlying this type of analysis is the representation of a

---

*The ordinate is 20 times the logarithm to the base 10 of the magnitude of the function (commonly called decibels although the function may not be associated directly with power), and the abscissa is the phase angle.

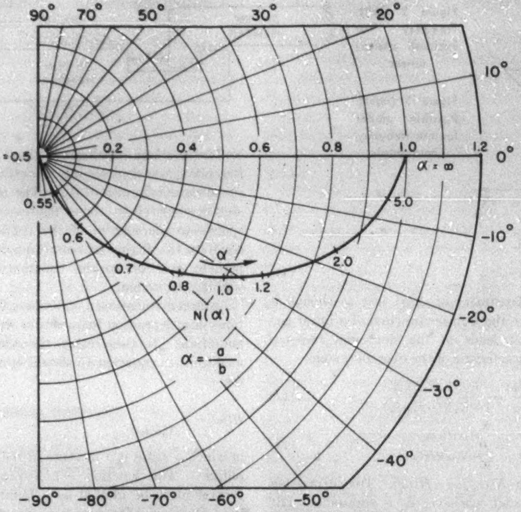


Figure 5. N(α), the free-play describing function

nonlinear element by some device—namely, a describing function—which is easier to deal with than the original nonlinear relationship. Simplification results from certain assumptions concerning the nature of the driving force imposed upon the element. These assumptions in turn place some limitations on the validity of the analysis. Thus simplification is obtained only at the expense of generality.

More specifically, in the sinusoidal analysis with which this paper is concerned, describing functions are calculated on the basis that the nonlinear elements have sinusoidal drives. These describing functions are applicable, therefore, when the inputs to the nonlinear elements are nearly sinusoidal with amplitude and frequency changing slowly enough that the elements are operating essentially in the steady state.

Describing functions in general still are nonlinear functions. They are slowly varying when valid, however, in contrast to the original relationships which may change rapidly and over wide ranges during a cycle. Use of describing-function representation of nonlinear elements substitutes a quasi-linear system of equations whose parameters change slowly if at all for one containing rapidly varying relationships.

The new system with slowly changing parameters can be thought of in terms of the poles and zeros of its response functions. As amplitude and frequency shift, the poles and zeros of the hypothetical system move about in the complex plane. In diagrams such as Figures 6 and 7, the $-1/N(\alpha)$ locus, by virtue of equation 15, can be considered to be the locus of the critical point, which in conventional theory is the -1 or -180-degree 0-decibel point. When the critical point lies to the left of the $H(i\omega)$ locus, the poles of the hypothetical system have negative real parts, and the system is said to be stable at this operating condition in the sense that any disturbances which appear tend to die out. On the other hand, when the critical point lies to the right of $H(i\omega)$, the poles have positive real parts and operation is unstable. If the $H(i\omega)$ locus passes through the critical point, as at the intersections in Figures 6 and 7, the real parts are zero, indicating a periodic solution.

Such a periodic solution may be stable or unstable with respect to a slight disturbance in the amplitude of oscillation. If the solution is stable, the amplitude tends to return to its original value following a disturbance; if unstable, the oscillation either increases or decreases indefinitely depending on the direction of the disturbance. Suppose that the system whose loci are shown in Figure 7 is operating at the amplitude corresponding to the intersection. Now assume that the amplitude is somehow given a slight increase so that the operating point on $-1/N(\alpha)$ shifts to the left or stable side of $H(i\omega)$. The amplitude consequently tends to decrease. As it does, the operating point moves back to its original position at the intersection. On the other hand, if the amplitude were to be decreased slightly, the operating point would shift to the right side of $H(i\omega)$ where poles of the

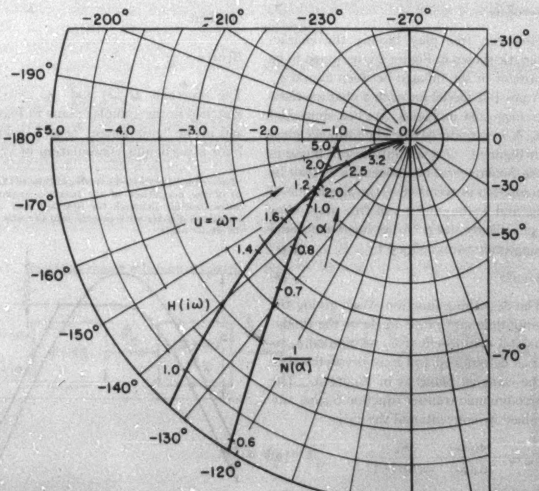


Figure 6. H(iω) and − 1/N(α) in the complex plane

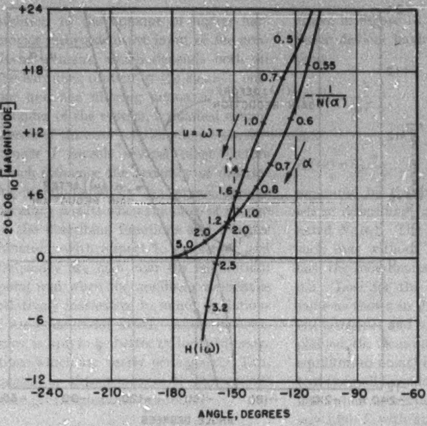


Figure 7. H(iω) and −1/N(α) in the gain-phase plane

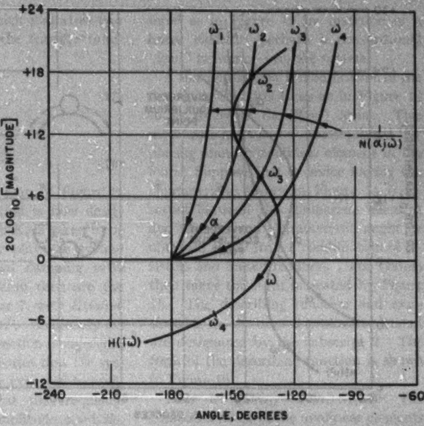


Figure 8. Example showing a describing function depending on frequency and amplitude

hypothetical system have positive real parts. The tendency of the amplitude to increase in this case would again return the operating point to its initial position at the intersection. The periodic solution indicated by the intersection in Figures 6 and 7, therefore, is a stable one. The intersection itself is termed a stable or convergent equilibrium point. The amplitude of the periodic solution is analogous to the position of a ball lying in the bottom of a hollow: the ball always rolls back to its original position after a disturbance.

Figure 9 shows a possible locus configuration in which there are two intersections. Intersection 1 is a convergent equilibrium point of the type just discussed. Intersection 2, however, is an unstable or divergent equilibrium point. If the amplitude is decreased slightly from α_2, the operating point shifts to the stable side of $H(i\omega)$. The amplitude decreases further, and in the process the system becomes more stable. A slight decrease in amplitude thus initiates a cumulative process in which the oscillation rapidly dies out. Conversely, an increase in amplitude from that at α_2 moves the operating point to the unstable region on the right of $H(i\omega)$ with the result that the amplitude increases further. This likewise is a cumulative process. The situation resembles that of a ball resting exactly on the top of a smooth hill, ready to roll off at the least disturbance.

This discussion of stability applies to frequency-variant describing functions as well as those that depend on amplitude alone. Thus, in Figure 8, the intersection at ω_3 is a convergent equilibrium point and that at ω_2 a divergent equilibrium point.

A convergent equilibrium point represents a possible stable oscillation. Conventional feedback oscillators depend for their operation on the existence of such points. On the other hand, unstable periodic solutions are physically unrealizable because minute random disturbances always present in an actual system are sufficient to prevent the delicate balance required. In general, a system must be excited externally with a definite amount of energy in order to reach a divergent equilibrium point. Consequently, the system represented by the loci of Figure 9, if isolated with no initial stored energy, would appear to be stable. When excited or driven, however, the relative stability would decrease as the amplitude of excitation increased until, if given sufficient energy, the system would become unstable. The amplitude at the natural mode would then increase until the convergent equilibrium point was reached. The system would oscillate stably at this point in the absence of excitation. Very large excitations would drive the operating point toward the point (−180-degrees, 0) on $-1/N(\alpha)$ into the stable region. Upon reduction of the excitation, operation would most likely resume at the convergent equilibrium point. The system could be shocked back to its original low-energy state, however, below the divergent equilibrium point.

Actually, continuation of the loci of Figure 7 to the small-amplitude-low frequency region would reveal a second intersection, this one representing an unstable periodic solution. However, for practical control systems, this solution is of little significance since, if an effort is made to minimize the amount of free play, disturbances nearly always will be present of sufficient magnitude to cause operation on the large-amplitude side of the divergent equilibrium point. In fact, even in simulating the system electronically, small steady-state oscillations at the convergent equilibrium point were found difficult to avoid.

A system whose equilibrium points occur in the reverse sequence to those of Figure 9 would behave quite differently. With the convergent equilibrium point at the lower amplitude, the system would be unstable at very small amplitudes. Stable oscillation would result at the convergent equilibrium point, but if the system were excited sufficiently to push the operating point beyond the divergent equilibrium point, the amplitude would increase without limit.

When the describing function does not depend on frequency, as in Figure 9, divergent equilibrium points always separate convergent equilibrium points. In a sense, the divergent points act as barriers tending to confine operation within one of a number of distinct regions. Somewhere in each region thus defined lies a convergent equilibrium point at which the system can oscillate when unexcited. The particular one at which oscillation actually occurs depends upon the region in which

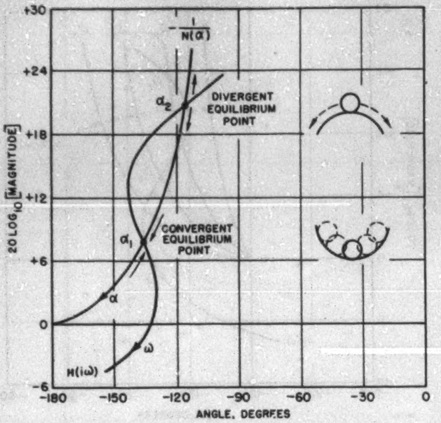


Figure 9. Convergent and divergent equilibrium points

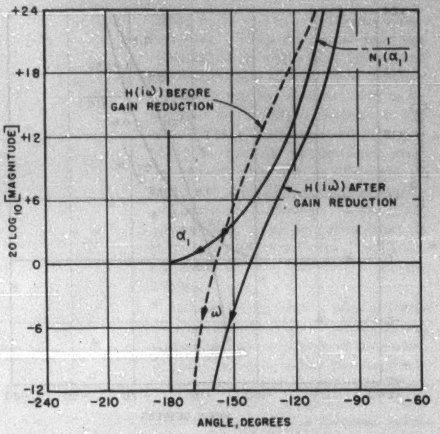


Figure 11. Loci for system of Figure 10, showing effect of decrease in loop gain

the excitation leaves the system. The equilibrium point at the smallest amplitude can be either divergent, as in Figure 9, or convergent. If divergent, the system is stable when isolated with low energy content. If convergent, the system tends to oscillate spontaneously with an amplitude corresponding to that at the convergent point. Both types of systems are common in feedback-control work.

Accuracy of Describing-Function Analysis

As already stated, the convergent equilibrium point present in Figures 6 and 7 indicates that a sustained oscillation can occur. The frequency and amplitude of the oscillation may be read directly as the values at the intersection of the loci. In this example, for instance, the predicted normalized frequency, $u=\omega T$, is 1.96 and the predicted normalized amplitude, $\alpha=a/b$, is 1.20. An exact analysis of the system shows that these values are in error by -0.97 per cent in frequency and -2.50 per cent in amplitude.

This discrepancy between the exact values and the results of the approximate analysis is due to the neglect of circulating harmonics in the system. The wave form of δ_a, Figure 4, obviously contains harmonics which are returned to the input of the nonlinear element through $H(i\omega)$ so that the initial assumption of a sinusoidal input is not strictly true. However, since $H(i\omega)$ provides some filtering action which attenuates the higher harmonics, the sinusoidal approximation is quite good.

Fortunately, most servo-type systems contain elements such as $H(i\omega)$ in the above example which act as low-pass filters. Consequently, even though the output of a nonlinear device be rich in harmonics, in a complete trip around the loop the harmonics may be so attenuated that the input to the nonlinear element is again nearly sinusoidal. This discrimination against higher harmonics tends to increase with the complexity or order of the system. In this lies one of the great attractions of the method since it means that more complicated systems, for which more conventional methods might be completely impractical, are just as effectively handled by describing-function analysis as are simple ones. Higher-order systems are more difficult to deal with only to the extent that their transfer functions are more difficult to plot. In this connection, it might be pointed out that the pertinent loci, representing both linear and nonlinear elements, might be obtained experimentally by direct measurements on the elements concerned rather than analytically as in the preceding example.

Clear-cut statements as to the types of nonlinear elements and systems for which the method is valid are difficult to make. The fundamental requirement for the nonlinear elements is that a sinusoidal input or driving force produce a response which contains a significant harmonic component having the same frequency as the input sinusoid. Although this restriction may appear to be trivial, it does eliminate some types of elements under certain conditions. A multivibrator, for instance, may be synchronized to a sinusoid having a frequency which is a submultiple of the running frequency. With the synchronizing signal considered as the input, the output does not contain a component of input frequency. In the analysis presented in Appendix I, the somewhat more stringent assumption is made that the output or response of a nonlinear element is periodic, having the same period as the input sinusoid. Thus an element producing a subharmonic response is excluded even though it may meet the first requirement mentioned.

Except for this mild but definite limitation, applicability of the method is generally a matter of degree, depending upon the nature of the linear portion of the system as well as the nonlinear elements. In

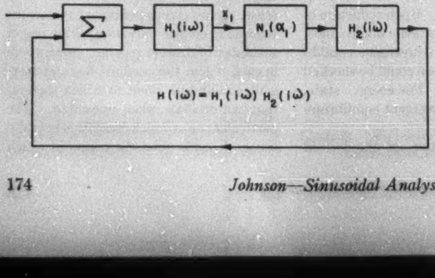


Figure 10. A control system containing free play

addition to the amount of higher harmonics returned to the input of the nonlinear element, which depends both on the harmonic content of the element output and the filtering action of the remainder of the system, a detailed examination of the error relations derived in Appendix I reveals several other factors which influence the accuracy of describing-function analysis. In general, greater accuracy results when the rates of change of the describing functions and transfer functions with respect to amplitude and frequency are high near the equilibrium point and when the nonlinear element is relatively insensitive to minor deviations from a sinusoidal input. Also, the accuracy is apt to be better at locus intersections which are nearly orthogonal. Tangential or almost tangential intersections may give misleading results.

Control and Prevention of Oscillations

The usefulness of sinusoidal or "frequency-response" analysis in the design of linear control systems is due primarily to the ease with which the effects of system modifications and additions can be evaluated. Thus, the shaping of transfer-function loci so as to obtain improved performance is a well-known and useful procedure. Extension of sinusoidal analysis through describing-function techniques carries over many of these same advantages to systems containing nonlinear elements.

As an example, consider the system drawn in Figure 10, which comprises two linear devices having the transfer functions

$$H_1(i\omega)=\frac{1}{i\omega T+1} \tag{17}$$

and

$$H_2(i\omega)=K_2/i\omega \tag{18}$$

separated by the free play of Figure 2, whose describing function is now designated $N_1(\alpha_1)$. The gain K_2 is taken to be such that without free play the system has the nondimensional damping ratio 0.2. Loci for the system then are the same as shown in Figure 7, with $H(i\omega)=H_1(i\omega)H_2(i\omega)$ and $u=\omega T$. As already explained, the locus intersection (convergent equilibrium point) indicates that the system normally would be expected to oscillate when unexcited at the frequency $\omega_0=1.96/T$ with an amplitude $a_1=1.2b_1$. The quantity a_1 is the amplitude of x_1, and b_1 is the amount of free play.

The steady-state oscillation may be undesirable for a number of reasons. It may cause excessive wear of gears and other mechanical parts; it may produce intolerable errors; or, where energy must be conserved as in a missile, the energy wasted may be unnecessarily high. The locus intersection evidently can be avoided by modifying either the $H(i\omega)$ locus or the $-1/N_1(\alpha_1)$ locus. Probably the simplest effective modification is a decrease in the system gain by a factor of about 3. The locus configurations then would be as sketched in Figure 11. Alternatively, the $H(i\omega)$ locus might be altered as in Figure 12 by insertion of a linear element—perhaps a conventional "lead" network—into the system.

A third possibility is modification of the describing-function locus as in Figure 13 so as to avoid the $H(i\omega)$ locus. This modification can be accomplished by placing another nonlinear element in the loop. Suppose that a device having the characteristic shown in Figure 14 is inserted between the comparison element and H_1. Such a characteristic might be obtained easily from a combination of resistors and biased rectifiers. The system then takes the form indicated by Figure 15. The describing function and associated α for the second nonlinear element are designated by the subscript 2. The form of the describing function is shown in Figure 16.

Now let $N(\alpha_1)=N_1(\alpha_1)N_2(\alpha_2)$. The combined effect of the nonlinear elements is studied by plotting the locus $-1/N(\alpha_1)$. In order to combine properly $N_1(\alpha_1)$ and $N_2(\alpha_2)$, α_2 must be found in terms of α_1. With

$$x_1=a_1\cos(\omega t+\theta_1) \tag{19}$$

and

$$x_2=a_2\cos(\omega t+\theta_2) \tag{20}$$

the amplitude a_1 is

$$a_1=|N_2(\alpha_2)H_1(i\omega)|a_2 \tag{21}$$

Since

$$a_1=\alpha_1 b_1 \tag{22}$$

and

$$a_2=\alpha_2 b_2 \tag{23}$$

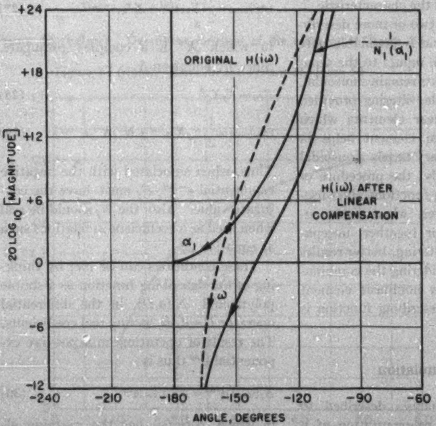


Figure 12. Effect of linear compensation on the loci for system of Figure 10

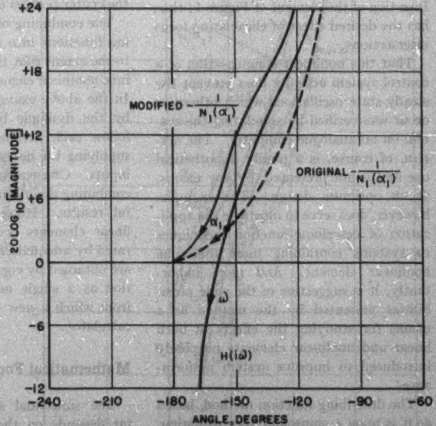
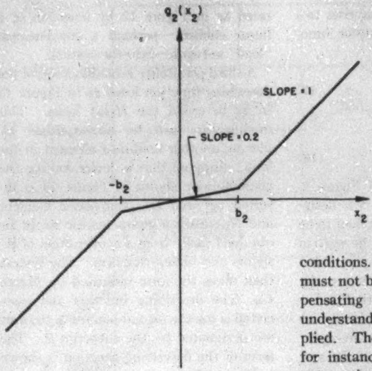


Figure 13. Effect of nonlinear compensation on the loci for system of Figure 10



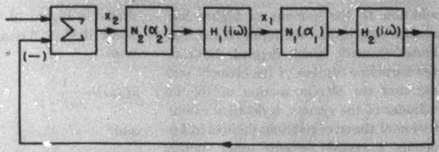


Figure 14 (left). Characteristic of nonlinear compensating element

Figure 15 (above). System of Figure 10 with nonlinear compensating element

equation 21 can be written

$$\alpha_1 = \alpha_2 |N_2(\alpha_2)H_1(i\omega)|(b_2/b_1) \quad (24)$$

Thus, for a given frequency and ratio of b_2/b_1, α_1 can be determined from a choice of α_2.

With $\omega=\omega_0$ (the frequency of oscillation of the original system) and $b_2/b_1=3.33$, the $-1/N$ locus has the form shown by the solid line in Figure 17. Loci of $-1/N$ at $\omega=0$ and $\omega=\infty$ are indicated by broken lines. At frequencies higher than ω_0, $-1/N$ lies in region 1 between $-1/N$ at ω_0 and the original $-1/N$ locus. For frequencies lower than ω_0, $-1/N$ lies to the left of $-1/N$ at ω_0 in region 2. Since $H(i\omega)$ enters region 1 only for $\omega<\omega_0$, a significant intersection of $H(i\omega)$ and $-1/N$ cannot occur at any frequency. Insertion of the element of Figure 14 thus has the desired effect of eliminating locus intersections.

That this nonlinear compensation of a control system actually does prevent the steady-state oscillations which otherwise occur was verified by simulating the system on an analogue computer. The system, of course, is a purely hypothetical one and is not presented, by any means, as an optimum design. The example, however, does serve to illustrate the application of describing-function techniques to systems containing more than one nonlinear element. And more importantly, it is suggestive of the wide possibilities presented by the method as a means for studying the effects of both linear and nonlinear elements purposely introduced to improve system performance.

The describing-function method, based as it is upon a sinusoidal approximation, obviously cannot give complete information regarding system behavior under all conditions. In view of this, the method must not be used blindly in adding compensating devices without a physical understanding of the consequences implied. The characteristic of Figure 14, for instance, amounts essentially to an automatic decrease in system gain for small errors. Possible detrimental effects of this gain reduction on other aspects of system performance should be examined. In this case there may be no such harmful consequences because at amplitudes for which the gain reduction is most effective, system performance would be ruined anyhow by the presence of the oscillation.

As already mentioned, shapes of the loci of the system of Figure 10 indicate that a reduction in system gain will prevent oscillation. Use of the nonlinear compensating element on this basis alone, therefore, without recourse to the analysis described above, is fairly reasonable. However, including the describing function of the Figure 14 element in a locus study provides a means for determining analytically a suitable slope and width for the center portion of the characteristic.

The combining of two or more describing functions into one is justifiable only to the extent that the inputs to the separate nonlinear elements remain sinusoidal. In the above example, filtering provided by the dynamic linear elements which follow each nonlinear element helps in supplying the necessary nearly sinusoidal inputs. Consequently, the procedure of combining describing functions gives useful results. However, when two nonlinear elements occur together, unseparated by adequate filtering, better results are obtained by considering the combination as a single new nonlinear element from which a new describing function is calculated.

Mathematical Formulation

The sinusoidal analysis described so far depends on the representation of a nonlinear element by its describing function. Since the describing function is concerned only with the fundamental component of the signal produced by the element, this representation obviously is incomplete. For the sake of simplicity initially, suppose that an element $g_j(x_j)$ is single valued and independent of time so that its describing function N_j is real. The describing-function representation of the element then is N_jx_j. The error introduced in replacing the true nonlinear relationship by N_jx_j is

$$g_j(x_j) - N_jx_j = \phi_j(x_j) \quad (25)$$

In terms of the describing function and the representation error $\phi_j(x_j)$, equation 1 may be written without loss in exactness as

$$x_k - N_jx_j = \phi_j(x_j) \quad (26)$$

This manipulation is straightforward and involves no difficulty for real N_j. When N_j is complex, however, additional interpretation is necessary if N_jx_j is to be the desired approximation of $g_j(x_j)$. From equations 3 and 4, the fundamental component of x_k is

$$(x_k)_{\text{fund}} = \frac{1}{2}[X_{k1}\epsilon^{i\omega t} + X^*_{k1}\epsilon^{-i\omega t}] \quad (27)$$

in which X^* is a complex conjugate. Since by equation 5

$$X_{k1} = N_jX_j \quad (28)$$

$$(x_k)_{\text{fund}} = \frac{1}{2}[N_jX_j\epsilon^{i\omega t} + N^*_jX^*_j\epsilon^{-i\omega t}] \quad (29)$$

Thus, when associated with the negative exponential $\epsilon^{-i\omega t}$, N_j must have the conjugate value. Also, the N_j should be real when used as a coefficient in relations such as equation 25.

These conditions can be met by thinking of the describing function as a simple polynomial, $N_j(a,D)$, in the differential operator $D=d/dt$, having real coefficients. The result of operating on a positive exponential $\epsilon^{i\omega t}$ thus is

$$N_j(a,D)\epsilon^{i\omega t} = N_j(a,i\omega)\epsilon^{i\omega t} \quad (30)$$

in which $N_j(a,i\omega)$ has the meaning already attached to the describing function.

The result of operating on a negative exponential is

$$N_j(a,D)\epsilon^{-i\omega t}\begin{aligned}&=N_j(a,-i\omega)\epsilon^{-i\omega t}\\&=N^*_{j*}(a^*,i^*\omega^*)\epsilon^{-i\omega t}\end{aligned} \qquad (31)$$

The N_j, when understood in this extended sense, evidently fulfill the requirements outlined here. The exact nature of $N_j(a,D)$ is unimportant since, with equation 25 regarded as the definition of $\phi_j(x_j)$, equation 26 is exact regardless. Furthermore, the functions on which $N_j(a,D)$ must operate in this analysis are periodic or almost periodic so that D is almost $i\omega$. Consequently, the essential property of $N_j(a,D)$ is simply that it coincide with $N_j(a,i\omega)$ on the substitution of $i\omega$ for D.

The linear, time-invariant elements of feedback-control systems usually can be described by ordinary linear differential equations having real constant coefficients. When this is the case, the most general relationships needed are of the form

$$\sum_{k=1}^{n} h_k(D)x_k=0 \qquad (32)$$

in which each $h_k(D)$ is a polynomial in the operator $D=d/dt$ having real constant coefficients. The x_k are the system variables, which are real functions of time. When nonlinear elements, or other types not describable in this manner, are present in a system, relationships such as equation 1 are needed also. A system containing several nonlinear elements then may be described mathematically by the following set of n equations:

$$\sum_{k=1}^{n} h_{jk}(D)x_k=g_j(x_j) \quad (j=1,2,\ldots n) \qquad (33)$$

The left sides of these equations represent the portions of the system which are considered to be linear in an engineering sense; the right side contains the remainder.

Equation 33 describes the force-free or unexcited system. If external driving functions are to be included, they may be added to the right side of the equations. Since describing functions have the nature of a transfer function between two quantities, they are useful directly only in dealing with nonlinear elements having but one input. Consequently, only such elements are included in the formulation of equation 33.

By means of relationships such as equation 26 rather than equation 1, the system may be formulated alternatively in terms of describing functions and representation errors as

$$\sum_{k=1}^{n} f_{jk}(D)x_k=\phi_j(x_j) \quad (j=1,2,\ldots n) \qquad (34)$$

The $f_{jk}(D)$ now contain the describing functions $N_j(a,D)$, and the representation errors $\phi_j(x_j)$ are defined by

$$\phi_j(x_j)=g_j(x_j)-N_j(a,D)x_j \qquad (35)$$

Equation 34 represents a very general situation and, as might be expected, is difficult to solve. One possible approach for finding periodic solutions is the perturbation procedure presented in Appendix I. For this purpose an additional parameter, μ, is inserted in equation 34 as a multiplier of the representation errors, ϕ_j

$$\sum_{k=1}^{n} f_{jk}(D)x_k=\mu\phi_j(x_j) \quad (j=1,2,\ldots n) \qquad (36)$$

Equation 36 now describes a hypothetical system. This system, however, coincides with the original one at $\mu=1$.

The solution to equation 36 is sought in the form of a power-series expansion in the parameter μ

$$x_k=x_{k0}+\frac{x_{k1}}{1!}\mu+\frac{x_{k2}}{2!}\mu^2+\ldots \qquad (37)$$

The frequency of the periodic solution, since it is unknown initially, is developed simultaneously as a similar power series

$$\omega=(1+\lambda)\omega_0 \qquad (38)$$

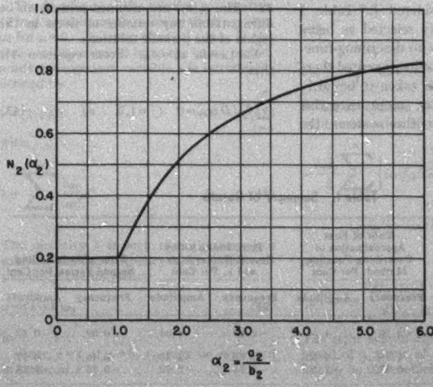
in which

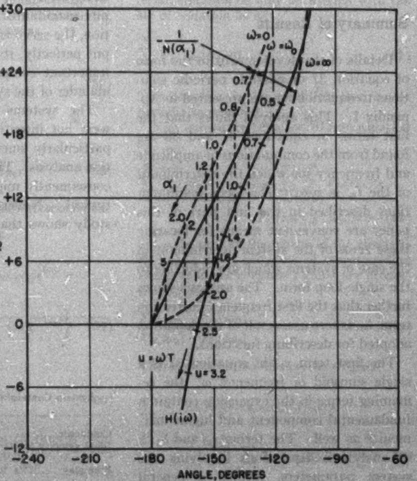
$$\lambda=\frac{\lambda_1}{1!}\mu+\frac{\lambda_2}{2!}\mu^2+\ldots \qquad (39)$$

Each of the coefficients x_{km} in equation 37 is a function of time, whereas the λ_m in equation 39 are constants. When the solution has been found in this form μ may be set equal to one to obtain the solution to the physical problem. This form of solution is quite appropriate for the purpose at hand because, as Appendix I shows, in the periodic case the representation of nonlinear elements by their describing functions leads naturally to the terms x_{k0} and ω_0 in equations 37 and 38. The first term in λ is zero. The next terms, x_{k1} and $\lambda_2/2$, provide a higher approximation to the solution which, if the series converges sufficiently rapidly, may be taken as an indication of the error of the analysis.

Figure 16 (below). Describing function of the element of Figure 14

Figure 17 (right). Loci for the system of Figure 15





Actual construction of equations 37 and 39 from the differential equations, as carried out in Appendix I, is a rather complicated process. In many respects it is similar to the usual power-series expansion of a function except that instead of the function itself being given, only its defining equations are known. For example, if x_k were available as a function of μ, the coefficients x_{km} in equation 37 might be found by differentiating x_k partially with respect to μ the appropriate number of times and evaluating the derivative at $\mu=0$. Thus

$$x_{km}=\left[\frac{\partial^m x_k}{\partial\mu^m}\right]_{\mu=0} \tag{40}$$

In the present case, x_k is unknown to begin with, but the defining differential equations are available. A similar procedure carried out on the differential equations themselves—differentiating partially with respect to μ and evaluating at $\mu=0$—yields successive sets of linear differential equations which may be solved by standard methods. The first set is homogeneous; the remainder have driving functions determined from solutions to the preceding sets.[6]

The process just described is straightforward provided the functions $\phi_j(x_j)$ are sufficiently regular that they can be expanded in power series of the x_j about a sinusoidal generating solution. Even though this requirement is not satisfied, however, higher approximations often can be developed by methods suggested by the work with better-behaved functions. This idea is discussed further in Appendix I.

Summary of Results

Details of the development in the form of equation 37 of possible periodic solutions to equation 34 are presented in Appendix I. This analysis shows that the first terms in equations 37 and 38 are found from the combinations of amplitude and frequency for which the determinant of the f_{jk} is zero. The locus manipulations described in the first part of this paper are convenient means for locating these zeros of the system determinant in the case of systems which are reducible to the single-loop form. The analysis shows further that the first frequency correction term, λ_1, is zero as a result of the definition adopted for describing functions.

The first term x_{k0} in equation 37 is a single sinusoid of frequency ω; the remaining terms in the expansion contain a fundamental component and higher harmonics as well. The terms x_{k1} and $\lambda_2/2$, as analytical expressions[3] involving the system parameters, furnish a general basis for quantitatively evaluating the effects of certain properties of the system on the accuracy of the describing-function method. These terms may be useful also either in estimating the error of the analysis in a given situation or as a means for achieving greater accuracy. The terms x_{k1} and $\lambda_2/2$ may be found in the general case in which the ϕ_j are analytic from equations 84 and 85. This expression is specialized in equation 99 to the single-loop case. Equation 101 is an approximation valid for nonanalytic ϕ_j.

In the example of the system containing free play discussed earlier, the errors of the describing-function analysis are stated to be −0.97 per cent in frequency and −2.50 per cent in amplitude. As an example of the applicability of x_{k1} and $\lambda_2/2$ as measures of the error of the analysis, their calculation from equation 101 predicts errors of −0.74 per cent and −2.22 per cent in frequency and amplitude respectively. These values compare favorably with the actual errors quoted previously. Put another way, the use of x_{k1} and $\lambda_2/2$ together with x_{k0} and ω_0 give the frequency and amplitude of the periodic solution to within −0.23 per cent and −0.28 per cent respectively.

These figures are tabulated in Table I along with comparable figures for two other systems for which the calculations have been made. The first system is a third-order system containing an ideal contactor (no deadspace and no hysteresis). The next is a second-order control system containing a servo which is velocity limited in the sense that the magnitude of its velocity is not able to exceed a definite maximum. Except for this limitation, the servo is assumed to follow its input perfectly, its dynamic lags being unimportant compared to those of the remainder of the system.

The systems considered in Table I were not intentionally selected as being particularly amenable to describing-function analysis. The results presented there consequently might be taken to be fairly typical. Nevertheless, more extensive study shows that for other systems, the accuracy may be considerably poorer. Indeed in some cases the describing-function method may fail completely to indicate existing periodic solutions. These failures in general occur when the H and $-1/N$ loci in diagrams such as Figure 7, for instance, approach each other but do not actually intersect. Good solid intersections, however, appear to produce first approximations which for a wide variety of systems are accurate to within 5 to 10 per cent. These results may then be further improved by roughly an order of magnitude through use of the next terms, x_{k1} and $\lambda_2/2$.

Appendix I

The Generating Solution

Equation 36, repeated here for convenience

$$\sum_{k=1}^{n} f_{jk}(D)x_k=\mu\phi_j(x_j) \quad (j=1,2\ldots n) \tag{41}$$

is presented under the section entitled "Mathematical Formulation" as the characterization in terms of describing functions of a system containing n nonlinear elements. As stated, a periodic solution to equation 41 is sought as a power-series expansion, equation 37, in the parameter μ

$$x_k=\sum_{m=0}^{\infty}\frac{x_{km}}{m!}\mu^m \tag{42}$$

Periodic solutions in this form have been developed by numerous workers, from Poincare on, for systems of varying complexity. Some work of Bulgakov is particularly applicable since it deals with systems of the desired generality.[4] Considerable use is made of this work in the following development.

Equation 41 is assumed to possess, under conditions that are to be investigated, solutions which are periodic in time and which can be represented in the form of equation 42. The $\phi_j(x_j)$ are assumed initially to be differentiable any number of times in the region of the periodic solutions.

For $\mu=0$, $x_k=x_{k0}$. From equation 41, then

$$\sum_{k=1}^{n} f_{jk}(D)x_{k0}=0 \quad (j=1,2,\ldots n) \tag{43}$$

Table I. Summary of Results

System Containing	Error of First Approximation or Describing Function Method, Per Cent		Error Indicated by Second Terms in x_k and λ, Per Cent		Error after Including Second Terms, Per Cent	
	Frequency	Amplitude	Frequency	Amplitude	Frequency	Amplitude
Contactor	−2.20	4.52	−2.29	4.94	0.09	−0.42
Ideal velocity limited servo	1.79	−4.53	1.65	−4.36	0.14	−0.17
Free play	−0.97	−2.50	−0.74	−2.22	−0.23	−0.28

In developing a solution in the form of equation 42, each x_{km} must be periodic if the complete solution is to be periodic over a finite range of μ. Consequently, it is assumed that for certain amplitudes of sinusoidal x_{k0}, the describing functions contained in the $f_{jk}(D)$ have values such that the determinant $F(D)$ of the elements $f_{jk}(D)$ has the simple pure imaginary zeros $\pm i\omega_0$. With the describing functions considered to be constants at just these values, equation 43 is a set of homogeneous linear differential equations with constant coefficients having a periodic solution which may be written

$$x_{k0}=\frac{a_0}{2}[F_{rk}(i\omega_0)\epsilon^{i\omega_0 t}+F_{rk}(-i\omega_0)\epsilon^{-i\omega_0 t}] \qquad (44)$$

$$=a_0 Re F_{rk}(i\omega_0)\epsilon^{i\omega_0 t} \qquad (45)$$

In these expressions $F_{rk}(i\omega_0)$ is the cofactor of $f_{rk}(i\omega_0)$ in the determinant $F(i\omega_0)$. The subscript r designates some row such that $F_{rs}(i\omega_0)$ is not zero for some column s.

The quantity a_0, which ordinarily is arbitrary in the solution of such a set of homogeneous equations, is fixed in this instance by the requirement that $F(\pm i\omega_0)=0$ when the amplitude in the kth describing function is $a_0|F_{rk}(i\omega_0)|$. In some cases there may be several sets of a_0 and ω_0 for which the determinant is zero. In other cases, of course, there may be no such values. This lack of zeros simply means that a periodic solution of this type does not exist. The amplitude and frequency, a_0 and ω_0, of x_{k0} may be found easily in simple single-loop systems as the values at the intersection of appropriate loci. This procedure is illustrated in the main body of the paper.

Throughout this development the $f_{jk}(D)$ are considered to be polynomials in D with real constant coefficients. This interpretation is possible because, as previously explained, the describing functions contained in the $f_{jk}(D)$ are evaluated at the particular value, a_0, of a that makes $F(i\omega_0)=0$. The remaining system nonlinearity is absorbed by the ϕ_j which now are

$$\phi_j(x_j)=g_j(x_j)-N_j(a_0,D)x_j \qquad (46)$$

The notation $N_j(a_0,D)$ indicates that the describing functions are evaluated at the generating solution amplitudes, $a_0|F_{rj}(i\omega_0)|$.

The frequency of the solution to equation 41 in general is not the same as the frequency ω_0 for $\mu=0$. In order that all terms of the series in equation 42 have the same period, a new independent variable τ is introduced, defined by

$$\tau=(1+\lambda)t \qquad (47)$$

with

$$\lambda=\sum_{m=1}^{\infty}\frac{\lambda_m}{m!}\mu^m \qquad (48)$$

The quantity λ is such that the complete solution has the frequency ω_0 in τ, and consequently the frequency

$$\omega=(1+\lambda)\omega_0 \qquad (49)$$

in t.

Since $\tau=t$ at $\mu=0$, x_{k0} may be expressed in terms of τ

$$x_{k0}=\frac{a_0}{2}[F_{rk}(i\omega_0)\epsilon^{i\omega_0\tau}+F_{rk}(-i\omega_0)\epsilon^{-i\omega_0\tau}] \qquad (50)$$

$$=a_0 Re F_{rk}(i\omega_0)\epsilon^{i\omega_0\tau} \qquad (51)$$

From equation 47

$$D=(1+\lambda)D' \qquad (52)$$

where D' is the operator

$$D'=d/d\tau \qquad (53)$$

At $\mu=0$, D evidently may be replaced by D'. Equation 51 therefore might have been obtained directly from equation 41 by setting $\mu=0$ and making the appropriate substitutions

$$\sum_{k=1}^{n} f_{jk}(D')x_{k0}=0 \quad (j=1,2,\ldots n) \qquad (54)$$

Equation 51 is a solution to this set of equations.

To find the second term in the series representation of x_k, equation 41 is differentiated with respect to μ

$$\sum_{k=1}^{n}\left\{f_{jk}(D)\frac{\partial x_k}{\partial\mu}+\frac{df_{jk}(D)}{dD}\frac{\partial D}{\partial\mu}x_k\right\}$$

$$=\phi_j(x_j)+\mu\frac{\partial\phi_j(x_j)}{\partial x_j}\frac{\partial x_j}{\partial\mu} \quad (j=1,2,\ldots n) \qquad (55)$$

Evaluation of this expression at $\mu=0$ results in

$$\sum_{k=1}^{n} f_{jk}(D')x_{k1}=\phi_j(x_{j0})-$$

$$\lambda_1 D'\sum_{k=1}^{n}\left[\frac{df_{jk}(D)}{dD}\right]_{D=D'}x_{k0} \qquad (56)$$

Since x_{j0} is sinusoidal with frequency ω_0 in τ, by the assumption on the nature of the nonlinear elements the $\phi_j(x_{j0})$ are periodic functions having as their fundamental frequency ω_0 also. Consequently they may be represented by the Fourier series

$$\phi_j(x_{j0})=\frac{1}{2}\sum_{m=-\infty}^{\infty}\Phi_{jm}\epsilon^{im\omega_0\tau} \qquad (57)$$

with

$$\Phi_{jm}=\frac{1}{\pi}\int_{-\pi}^{\pi}\phi_j(x_{j0})\epsilon^{-im\omega_0\tau}d\omega_0\tau \qquad (58)$$

By use of equation 50 the second term on the right in equation 56 is

$$-\frac{\lambda_1 a_0}{2}\sum_{k=1}^{n}\left\{i\omega_0 f_{jk}'(i\omega_0)F_{rk}(i\omega_0)\epsilon^{i\omega_0\tau}-\right.$$

$$\left.i\omega_0 f_{jk}'(-i\omega_0)F_{rk}(-i\omega_0)\epsilon^{-i\omega_0\tau}\right\} \qquad (59)$$

where

$$f_{jk}'(i\omega_0)=\left[\frac{df_{jk}(D)}{dD}\right]_{D=i\omega_0} \qquad (60)$$

The right side of equation 56 thus is composed of a series of exponentials, $\epsilon^{\pm im\omega_0\tau}$. Let B_j be the coefficient of $\epsilon^{i\omega_0\tau}$

$$B_j=\frac{1}{2}\left\{\Phi_{j1}-i\lambda_1\omega_0 a_0\sum_{k=1}^{n}F_{rk}(i\omega_0)f_{jk}'(i\omega_0)\right\} \qquad (61)$$

Then equation 56 may be written

$$\sum_{k=1}^{n} f_{jk}(D')x_{k1}=\frac{\Phi_{j0}}{2}+B_j\epsilon^{i\omega_0\tau}+B^*_j\epsilon^{-i\omega_0\tau}+$$

$$\frac{1}{2}\sum_{m=2}^{\infty}\left\{\Phi_{jm}\epsilon^{im\omega_0\tau}+\Phi^*_{jm}\epsilon^{-im\omega_0\tau}\right\}$$

$$(j=1,2,\ldots n) \qquad (62)$$

In order that the solution of this equation by periodic, the particular integral due to the terms $B_j\epsilon^{i\omega_0\tau}$ and $B^*_j\epsilon^{-i\omega_0\tau}$ must be of the form $C_k\epsilon^{i\omega_0\tau}$ and $C^*_k\epsilon^{-i\omega_0\tau}$, respectively, and not of the form $\tau\epsilon^{\pm i\omega_0\tau}$. Consider

$$\sum_{k=1}^{n} f_{jk}(D')x_{k1}=B_j\epsilon^{i\omega_0\tau} \quad (j=1,2,\ldots n) \qquad (63)$$

and assume that x_{k1} is $C_k\epsilon^{i\omega_0\tau}$. Substitute this into equation 63 and drop the common factor $\epsilon^{i\omega_0\tau}$ since it is never zero

$$\sum_{k=1}^{n} f_{jk}(i\omega_0)C_k=B_j \quad (j=1,2,\ldots n) \qquad (64)$$

In view of the assumption that $\pm i\omega_0$ are simple zeros of $F(D)$, the matrix of the elements $f_{jk}(i\omega_0)$ is of rank $n-1$, hence

$$\sum_{j=1}^{n} F_{js}(i\omega_0)f_{jk}(i\omega_0)=0 \qquad (65)$$

for any k. Multiply this equation by C_k and sum over k

$$\sum_{k=1}^{n} C_k\sum_{j=1}^{n} F_{js}(i\omega_0)f_{jk}(i\omega_0)=0 \qquad (66)$$

Since the order of summation is immaterial, equation 66 may be written with the aid of equation 64 as

$$\sum_{j=1}^{n} F_{js}(i\omega_0)B_j=0 \qquad (67)$$

This equation is the condition that x_{k1} be periodic as required.

Now substitute equation 61 into equation 67

$$\sum_{j=1}^{n} F_{js}(i\omega_0)\Phi_{j1}-i\lambda_1\omega_0 a_0\sum_{j=1}^{n}\sum_{k=1}^{n} F_{js}(i\omega_0)\times$$

$$F_{rk}(i\omega_0)f_{jk}'(i\omega_0)=0 \qquad (68)$$

Since the rank of the matrix of the elements $f_{jk}(i\omega_0)$ is $n-1$, cofactors of corresponding elements in the columns s and k are proportional; that is

$$\frac{F_{jk}(i\omega_0)}{F_{js}(i\omega_0)}=\frac{F_{rk}(i\omega_0)}{F_{rs}(i\omega_0)} \qquad (69)$$

Therefore equation 68 may be written

$$\sum_{j=1}^{n} F_{js}(i\omega_0)\Phi_{j1}-i\lambda_1\omega_0 a_0 F_{rs}(i\omega_0)\times$$

$$\sum_{j=1}^{n}\sum_{k=1}^{n} F_{jk}(i\omega_0)f_{jk}'(i\omega_0)=0 \qquad (70)$$

The derivative of the determinant $F(D)$ is

$$\frac{dF(D)}{dD}=\sum_{j=1}^{n}\sum_{k=1}^{n}F_{jk}(D)\frac{df_{jk}(D)}{dD} \qquad (71)$$

Consequently, equation 70 reduces to

$$\sum_{j=1}^{n}F_{js}(i\omega_0)\Phi_{j1}-i\lambda_1\omega_0a_0F_{rs}(i\omega_0)F'(i\omega_0)=0 \qquad (72)$$

This in general is a complex equation which must be satisfied in order that x_{k1} be periodic. The conjugate relation would have been obtained by considering $\boldsymbol{B}^*_j\epsilon^{-i\omega_0\tau}$ in the right side of equation 63. It will be satisfied whenever equation 72 is met.

From equations 46 and 58

$$\Phi_{j1}=\frac{1}{\pi}\int_{-\pi}^{\pi}g_j(x_{j0})\epsilon^{-i\omega_0\tau}d\omega_0\tau-a_0F_{rj}(i\omega_0)N_j(a_0,i\omega_0) \qquad (73)$$

Evidently Φ_{j1} is zero at a_0 and ω_0 if

$$N_j(a_0,i\omega_0)=\frac{1}{a_0F_{rj}(i\omega_0)\pi}\int_{-\pi}^{\pi}g_j(x_{j0})\epsilon^{-i\omega_0\tau}d\omega_0\tau \qquad (74)$$

But equation 74 coincides with definition 5 of the describing function, since $a_0F_{rj}(i\omega_0)$ is the complex amplitude of the input x_{j0}. The adopted definition of describing function thus forces the Φ_{j1} to be zero and leads to the important result, from equation 72, that the first periodicity condition is satisfied when the first frequency correction is zero.

Equation 74, which requires that

$$\Phi_{j1}(a_0,i\omega_0)=0 \qquad (75)$$

might be regarded as an alternative definition of the describing function. From equation 5, the describing function at any amplitude and frequency is

$$N_j(a,i\omega)=\frac{1}{aF_{rj}(i\omega)\pi}\int_{-\pi}^{\pi}g_j(x_{j0})\epsilon^{-i\omega\tau}d\omega\tau \qquad (76)$$

in which now

$$x_{j0}=a\,Re\,F_{rj}(i\omega)\epsilon^{i\omega\tau} \qquad (77)$$

Consequently, for future use $\Phi_{j1}(a,i\omega)$ may be written

$$\Phi_{j1}(a,i\omega)=aF_{rj}(i\omega)[N_j(a,i\omega)-N_j(a_0,i\omega_0]\qquad (78)$$

Second Term in the Series Solution

Equations 62 are a set of inhomogeneous linear differential equations. The periodic solution may be written

$$x_{k1}=\frac{1}{2}\sum_{m=-\infty}^{\infty}Y_{km}\epsilon^{im\omega_0\tau}+b\,Re\,F_{rk}(i\omega_0)\epsilon^{i\omega_0\tau} \qquad (79)$$

in which

$$\begin{aligned}Y_{km}&=0 && (m=\pm1)\\ &=\sum_{j=1}^{n}\frac{\Phi_{jm}F_{jk}(im\omega_0)}{F(im\omega_0)} && (m\neq\pm1) \qquad (80)\end{aligned}$$

The quantity b contains an integration constant, so far undetermined. This term is present because, as assumed earlier, the system determinant is zero at $\pm i\omega_0$. The further assumption is made in writing equations 79 and 80 that the system determinant is not zero at multiples of $\pm i\omega_0$; that is, $F(im\omega_0)\neq0$ except for $m=\pm1$.

The second term in the series expansion for x_k thus contains the results of higher harmonics generated by the x_{k0} through the nonlinear functions ϕ_j, as well as an additional contribution to the fundamental component.

In general, each of the terms x_{km} contains a component of fundamental frequency. The first term x_{k0} is entirely fundamental. Its phase clearly is completely arbitrary, since any difference in phase amounts merely to a shift along the time axis. Bulgakov shows that in the higher terms the fundamental component in phase with x_{k0} is determined along with the terms in λ by successive periodicity conditions.[4] The out-of-phase, or quadrature, fundamental component, however, is arbitrary provided only that

$$\frac{c_1}{1!}\mu+\frac{c_2}{2!}\mu^2+\ldots \qquad (81)$$

converges, where c_m is the amplitude of the out-of-phase fundamental in x_{km}. An out-of-phase fundamental in one term affects the remaining terms in the series, of course, but it may be specified independently to meet some boundary condition. The requirement that all out-of-phase fundamental components be zero is the simplest such condition. Consequently, in equation 79 the increment in fundamental component of amplitude $b|F_{rk}(i\omega_0)|$ is taken to be in phase with x_{k0}.

Equation 41 now is differentiated a second time with respect to μ and evaluated at $\mu=0$.

The result, with $\lambda_1=0$, is

$$\sum_{k=1}^{n}f_{jk}(D')x_{k2}=2\frac{d\phi_j(x_{j0})}{dx_{j0}}x_{j1}-\lambda_2D'\sum_{k=1}^{n}\left[\frac{df_{jk}(D)}{dD}\right]_{D=D'}x_{k0} \qquad (j=1,2,\ldots n) \qquad (82)$$

As before, in order to avoid terms in x_{k2} of the form $\tau\epsilon^{\pm i\omega_0\tau}$

$$\sum_{j=1}^{n}F_{js}(i\omega_0)B_j=0 \qquad (67)$$

in which B_j is the coefficient of $\epsilon^{i\omega_0\tau}$ in the Fourier expansion of the right side of equation 82

$$B_j=\frac{1}{\pi}\int_{-\pi}^{\pi}\frac{d\phi_j(x_{j0})}{dx_{j0}}x_{j1}\epsilon^{-i\omega_0\tau}d\omega_0\tau-\frac{i\lambda_2\omega_0a_0}{2}\sum_{k=1}^{n}F_{rk}(i\omega_0)f_{jk}'(i\omega_0) \qquad (83)$$

Substitution of equation 83 into equation 67 leads to the second periodicity condition which after some manipulation can be written

$$a_0b\sum_{j=1}^{n}F_{rj}(i\omega_0)F_{js}(i\omega_0)\left[\frac{\partial N_j(a,i\omega_0)}{\partial_a}\right]_{a_0}-\frac{i}{2}\lambda_2\omega_0a_0F_{rs}(i\omega_0)F'(i\omega_0)=\sum_{j=1}^{n}F_{js}(i\omega_0)M_j \qquad (84)$$

with

$$M_j=-\frac{1}{2\pi}\sum_{m=-\infty}^{\infty}Y_{jm}\int_{-\pi}^{\pi}\frac{d\phi_j(x_{j0})}{dx_{j0}}\times\epsilon^{i(m-1)\omega_0\tau}d\omega_0\tau \qquad (85)$$

The two real quantities b and λ_2 are determined from the complex equation 85.

Convergence of the Series Solution

Development of the formal solution for x_k as begun in the preceding sections may be carried on indefinitely if the $\phi_j(x_j)$ are analytic in the region about the generating solution x_{j0}. This requirement implies that the functions $\phi_j(x_j)$ may be expanded in a power series in (x_j-x_{j0}) which converges for

$$|(x_j-x_{j0})|<\delta_j \qquad (86)$$

Furthermore, the formal solution as a power series in μ converges for μ small enough in terms of the magnitudes of the ϕ_j and the quantities δ_j of equation 86. Unfortunately, no way is available at the present time for finding exactly how large μ may be, still permitting convergence. The real topic of interest, however, is not simply whether or not equation 42 converges. The hoped-for situation is that the first terms in the x_k and ω series will be sufficiently accurate to give information useful for the engineering of control systems. Further, it is hoped that the second terms in x_k and λ, as analytical expressions involving the properties of the system, will provide a basis for discussing the applicability of the general describing-function method. Alternatively, the second terms might be used in the unusual problems requiring accuracy higher than that of the first terms. The important question, then, is: what is the accuracy of the first few terms of x_k and ω?

This question is hardly more amenable to neat mathematical treatment than is the question of convergence. The questions obviously are not the same, since the accuracy of the early terms can be quite satisfactory even though the series fail to converge. Conversely, the first terms may be completely useless in spite of eventual convergence of the series. The question of the preceding paragraph can be answered to some extent by investigation of specific examples. Thus the data presented in the main body of the paper indicate that the accuracy of the terms is, indeed, quite satisfactory in several typical situations.

Nonanalytic Elements

When the ϕ_j are not analytic functions, their derivatives in equation 85 have no meaning. Closer examination shows that M_j (equation 85) represents essentially the change in the fundamental component of $\phi_j(x_j)$ due to the higher harmonics of x_{j1}. Consequently, equation 84 still may be used

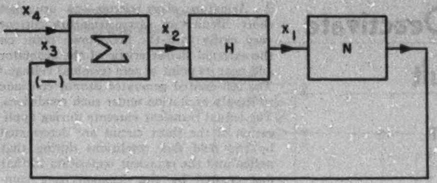


Figure 18. Single-loop system containing one nonlinear element

to find approximate amplitude and frequency corrections, b and $\lambda_2/2$, even though the functions involved are nonanalytic if equation 85 is replaced by

$$M_j = -\Delta\Phi_{j1}(a_0, i\omega_0) \qquad (87)$$

In this expression, $\Delta\Phi_{j1}(a_0, i\omega_0)$ designates the change in the complex amplitude of the fundamental component of $\phi_j(x_j)$ when the higher harmonic terms of equation 79 are added to equation 50. The derivative of N_j in equation 84 is well defined in nearly all cases of interest even though the ϕ_j are not analytic.

Single-Loop Systems Containing One Nonlinear Element

Analysis of the general situation, as outlined in the foregoing sections, is necessarily rather complex. Considerable simplification is possible when the system contains only one nonlinear element, as in Figure 18. The actual physical arrangement of the system need not be this simple. As far as the effect of the nonlinear element on stability is concerned, however, the single-nonlinearity system can always be placed in this form.

In Figure 18 the linear portions of the system are lumped into one linear element designated by the transfer function $H(i\omega)$. The element may also be specified by its differential equation

$$f_{21}(D)x_1 + f_{22}(D)x_2 = 0 \qquad (88)$$

In terms of this equation, the transfer function is

$$H(i\omega) = -\frac{f_{22}(i\omega)}{f_{21}(i\omega)} \qquad (89)$$

The output x_3 of the nonlinear element is some function of its input x_1

$$x_3 = g_1(x_1) \qquad (90)$$

This equation can be placed in the form of equation 26 by subtracting $N_1(a,D)x_1$ from both sides

$$-N_1(a,D)x_1 + x_3 = \phi_1(x_1) \qquad (91)$$

In this expression, N_1 is the describing function for the nonlinear element, and the function $\phi_1(x_1)$ is defined by

$$\phi_1(x_1) = g_1(x_1) - N_1(a,D)x_1 \qquad (92)$$

Since this analysis is concerned with the unexcited or force-free system, $x_3 = -x_2$. With this replacement and the insertion of μ as a multiplier of $\phi_1(x_1)$, equation 91 takes the form

$$-N_1(a,D)x_1 - x_2 = \mu\phi_1(x_1) \qquad (93)$$

The force-free system thus is described by equations 88 and 93. The determinant is

$$F(i\omega) = \begin{vmatrix} -N_1(a,i\omega) & -1 \\ f_{21}(i\omega) & f_{22}(i\omega) \end{vmatrix} \qquad (94)$$

$$= f_{21}(i\omega)[1 + N_1(a,i\omega)H(i\omega)] \qquad (95)$$

The amplitude a in $N_1(a,i\omega)$ now is fixed at some value a_0 such that $F(\pm i\omega_0) = 0$ for some frequency ω_0. The linear portions of the system are assumed to be stable so that $f_{21}(i\omega) \neq 0$. Consequently

$$1 + N_1(a_0, i\omega_0)H(i\omega_0) = 0 \qquad (96)$$

or

$$H(i\omega_0) = -\frac{1}{N_1(a_0, i\omega_0)} \qquad (97)$$

This is essentially the same condition as equation 15 given earlier. Let the subscripts r and s of the preceding analysis be 2 and 1 respectively.

Then

$$x_{10} = a_0 \cos \omega_0\tau \qquad (98)$$

After considerable manipulation, the relation between b and λ_2 for analytic $g_1(x_1)$, from equations 84 and 85, can be written

$$a_0 b\left[\frac{dN_1(a)}{da}\right]_{a_0} + \frac{i\lambda_2\omega_0 a_0 N_1(a_0)H'(i\omega_0)}{2H(i\omega_0)}$$
$$= -\frac{Y_{10}}{2}\left[\frac{dG_{10}}{da}\right]_{a_0} - \sum_{m=2}^{\infty}\left\{\left[\frac{dG_{1m}}{da}\right]_{a_0} \times \right.$$
$$\left. Re\,Y_{1m} + \frac{im}{a_0}G_{1m}\,Im\,Y_{1m}\right\} \qquad (99)$$

In this equation

$$G_{1m} = \frac{1}{\pi}\int_{-\pi}^{\pi} g_1(x_{10})\epsilon^{-im\omega_0\tau}\,d\omega_0\tau \qquad (100)$$

Also, analyticity of $g_1(x_1)$ implies that g_1 is a single-valued function not involving time or derivatives of x_1. Such functions lead to real describing functions which depend on amplitude only. Consequently the describing function in equation 99 is written $N_1(a)$.

Equation 99 may be valid even though the nonlinear element is not analytic provided that its characteristic can be construed to be a limiting form of an analytic relationship. For example, a contactor step function without hysteresis can be reached through a limiting process from a characteristic having rounded corners and finite slope through the origin. Consequently, equation 99 in general is applicable to single-valued characteristics.

When the element is not of this type or when equation 99 is inconvenient, the quantities b and λ_2 may be found from the following approximate relationship, for which equation 87 rather than equation 85 is used

$$a_0 b\left[\frac{\partial N_1(a, i\omega_0)}{\partial a}\right]_{a_0} + \frac{i\lambda_2\omega_0 a_0}{2H(i\omega_0)} \times$$
$$\left[\frac{d}{di\omega}N_1(a_0, i\omega)H(i\omega)\right]_{\omega_0}$$
$$= -\Delta G_{11}(a_0, i\omega_0) \qquad (101)$$

The increment ΔG_{11} has the same meaning with respect to g_1 as has $\Delta\Phi_{11}$ of equation 87 with respect to ϕ_1. In fact $\Delta G_{11} = \Delta\Phi_{11}$.

The frequency correction $\lambda_2/2$ is easily separated from b in equation 99. Thus,

$$\frac{\lambda_2}{2} = \frac{H(i\omega_0)}{\omega_0 a_0^2 N_1(a_0)\,Re\,H'(i\omega_0)}\sum_{m=2}^{\infty} mG_{1m}^2\,Im \times$$
$$\frac{H(im\omega_0)}{1 + N(a_0)H(im\omega_0)} \qquad (102)$$

From this equation can be derived some well-known relations of oscillator theory.

References

1. A FREQUENCY RESPONSE METHOD FOR ANALYZING AND SYNTHESIZING CONTACTOR SERVOMECHANISMS, **Ralph J. Kochenburger.** *AIEE Transactions*, volume 69, part I, 1950, pages 270–84.

2. INTRODUCTION TO NON-LINEAR MECHANICS (book), **N. Kryloff, N. Bogoliuboff,** free translation by **S. Lefschetz.** Princeton University Press, Princeton, N. J., 1943.

3. NON-LINEAR MECHANICS (book), **N. Minorsky.** J. W. Edwards, Ann Arbor, Mich., 1947.

4. PERIODIC PROCESSES IN FREE PSEUDO-LINEAR OSCILLATORY SYSTEMS, **B. V. Bulgakov.** *Journal*, Franklin Institute (Philadelphia, Pa.), volume 235, June 1943, pages 591–616.

5. CARRIER-CONTROLLED RELAY SERVOS, **J. C. Lozier.** *Electrical Engineering*, volume 69, number 12, December 1950, pages 1052–56.

6. FORCED OSCILLATIONS IN OSCILLATOR CIRCUITS AND THE SYNCHRONIZATION OF OSCILLATORS, **D. G. Tucker.** *Journal*, Institution of Electrical Engineers (London, England), volume 92, part III, 1945, pages 226–34.

7. THE EFFECTS OF BACKLASH AND OF SPEED DEPENDENT FRICTION ON THE STABILITY OF CLOSED-CYCLE CONTROL SYSTEMS, **A. Tustin.** *Journal*, Institution of Electrical Engineers (London, England), volume 94, part IIA, 1947, pages 143–51.

No Discussion

Short-Circuiting Techniques to Deactivate a Generator After a Fault

OSCAR MARKOWITZ
MEMBER AIEE

J. W. POBST
MEMBER AIEE

Synopsis: This paper presents a method for protecting an aircraft self-excited d-c generator after either a full-field short-circuit fault or a generator feeder fault has been detected. Protection is afforded, regardless of whether the fault occurs inside the machine or externally. It is also applicable to other types of faults. The method consists, essentially, of purposely applying short circuits between the different generator output terminals, thereby disabling the power-generating capabilities of the machine.

AIRCRAFT electric system protection has stimulated much thought and activity during recent years. As operating voltage and output power have increased, so have their interruption requirements. There has been a tendency to increase the size of the interrupting device and, recently, to provide some form of backup protection. The deficiencies of most of these types of circuit protection are rather well known. There have been cases of generator contactor failure caused by insufficient arc-interrupting capacity or contact welding. Even with normal contactor operation during faulted conditions, there have been no means to de-energize the generator completely for all types of faults. Those responsible for the aircraft electric systems generally have recognized the desirability of providing protection equipment for a self-excited d-c generator which, after a full-field short-circuiting fault, would reduce the generator terminal voltage to zero, or, after a generator feeder fault, would reduce the current output of the machine essentially to zero. Presently available equipment and circuits do not provide assurance of this type of protection.

No attempt has been made herein to discuss the sensing of the fault; this is considered an independent and adequate entity. Also, no attempt was made to use this short-circuiting technique with paralleled generators, since, if sensing has been correctly utilized, it may be assumed that the sensing device has detected the faulted condition and has selected the proper generator for deactivation.

Theory of Operation

Previous protection arrangements have been inadequate for the following types of faults:

1. Full-field short-circuiting fault—a full-field fault will allow complete shunt-field excitation after all control circuits have been opened.
2. Generator feeder fault—a generator feeder fault will permit heavy currents to feed the fault continuously when there is sufficient current in the series field to provide the required internal flux.

In both cases the power loops remain closed, even after disconnections have been made between the generator and the main bus, presenting a possible continuous hazard.

To control a generator after such faults, its internal magnetic flux should be reduced to the point where no electric power can be generated. The first attempt to reduce excitation during overvoltage was made by purposely connecting a short circuit across the shunt field to by-pass the field current. This idea of purposeful short circuits was then extended to the other generator terminals (or their extensions) singly and in combinations for more complete control. These short circuits, together with adequate circuit preparations, by-pass the currents that allow the continuation or secondary build-up of internal flux.

Four types of purposeful short circuits may be considered for control of the faulted system:

1. Shunt-field short circuit—to provide a shunt-field short circuit, the field circuit between the short circuit and the generator bus first must be open-circuited. The short circuit then will provide a by-pass for current which might flow through the shunt field during a full-field short-circuiting fault. During other types of faults, the short circuit will provide a field-energy discharge path, substantially helping other simultaneous actions in reducing terminal voltage. Such a short circuit, by itself, cannot reverse the residual flux of the generator because of inherent shunt-field resistance and hysteresis.

2. Armature short circuit—an armature short circuit can carry armature current very easily past the breakdown point on the external characteristic of the generator and near its point of zero terminal voltage.[1] The self-excited generator cannot continue to supply excitation under such conditions. The actual transient currents during application of the short circuit are determined by the field flux conditions during that period and the transient opposition to that flux created by the demagnetizing component of the armature magnetomotive force. If there is not sufficient original flux, the short-circuiting current, under certain conditions, can produce enough demagnetizing component together with the circuit inductivity to reverse the residual flux of the generator.

Before this armature short circuit can be applied, the circuit between the generator and other paralleled sources of power must be conditioned so that no power, or only a little predetermined amount of power, can be fed back to the short circuit from these other sources. This requirement can be met with a contactor which opens the circuit between the short circuit and the main bus. A resistor placed across the contacts of this contactor will decrease the current to a value which will allow, after completion of the short circuit, for the proper operation of the reverse-current relay normally used with the system.

3. Generator feeder-to-ground short circuit—A generator feeder-to-ground short circuit essentially accomplishes the effect of the armature short circuit and requires the same circuit preparation. The difference lies in that the short-circuiting currents now flow through the different series-connected windings of the machine as well as the armature itself. These windings limit the flux distortion caused by armature current and also contribute to the in-line shunt-field flux, thus completely avoiding the possibility of reversed residual flux.[2] This short circuit, alone, can permit the generator current to maintain itself above approximately mid-speed.

4. Negative-brush-to-ground short circuit—the generator equalizer (or paralleling) terminal, which is connected internally to the negative brush of the machine or to a point close to the negative brush, generally is brought out as an external terminal. A short circuit from this equalizer terminal to ground will by-pass an effective proportion of the current flowing through the series-connected windings. If this short circuit is accomplished after the generator feeder-to-ground short circuit has been completed, it will by-pass sufficient current to avoid current self-excitation of the machine so that the short circuits will not be fed with power after the initial transient. The equalizer bus should be disconnected properly from the generator during and after the control process to minimize voltage disturbances at the bus.

Test Methods

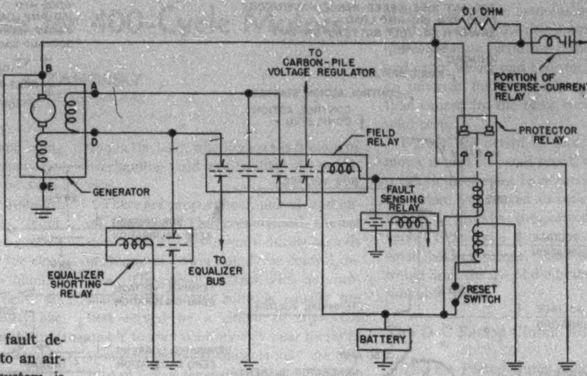
A 30-volt 300-ampere wide-speed-range d-c generator, with an appropriate regulator, reverse-current relay, and additional components to provide the various

---

Paper 52-139, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE South West District Meeting, St. Louis, Mo., April 15–17, 1952. Manuscript submitted April 6, 1951; made available for printing February 27, 1952.

OSCAR MARKOWITZ and J. W. POBST are both with the United States Naval Air Development Center, Johnsville, Pa.

Figure 1. Circuit used for generator control after a fault


0.1 OHM
TO MAIN BUS
PORTION OF REVERSE-CURRENT RELAY
PROTECTOR RELAY
TO CARBON-PILE VOLTAGE REGULATOR
FIELD RELAY
FAULT SENSING RELAY
B
A
D
E
GENERATOR
EQUALIZER SHORTING RELAY
TO EQUALIZER BUS
RESET SWITCH
BATTERY

purposeful short circuits after fault detection, furnished equivalence to an aircraft electric system. This system is shown in Figure 1. It was used throughout the tests for generator control after the faults had been sensed or the sensing had been simulated.

A 2-coil protector relay with heavy contacts and having a magnetic latch in its normally closed position was used for the purposeful generator feeder short circuits The equalizer-terminal-to-ground short circuit was accomplished by a normally closed relay which was de-energized by a reduction in generator terminal voltage. Opening and short-circuiting of the shunt-field circuit was achieved by a field relay with contacts as indicated in Figure 1. Thus, three separate relays were used to accomplish these purposeful short circuits. It is both possible and desirable, however, to combine their individual functions into the protector relay with proper time-sequencing of the contacts which perform the circuit preparations and purposeful short circuits.

Full-field short-circuiting faults were sensed by an overvoltage relay. Generator feeder faults and detection were simulated, with control occurring approximately 50 milliseconds after the fault was applied.

Results and Discussion

The sequence of circuit and relay operation once the fault has been sensed is indicated by the encircled numbers shown in Figure 2. The millisecond values specified are those considered applicable to good circuit and generator protection after a fault utilizing the techniques discussed.

Representative conditions encountered in the operation of an aircraft electric system were selected. Oscillograms of these conditions were taken in successive order and are shown in Figures 3, 4, and 5. Figure 3 shows the circuit condition obtained after a full-field fault. Values of 6,000 rpm and a 150-ampere load were selected to simulate cruising speed with

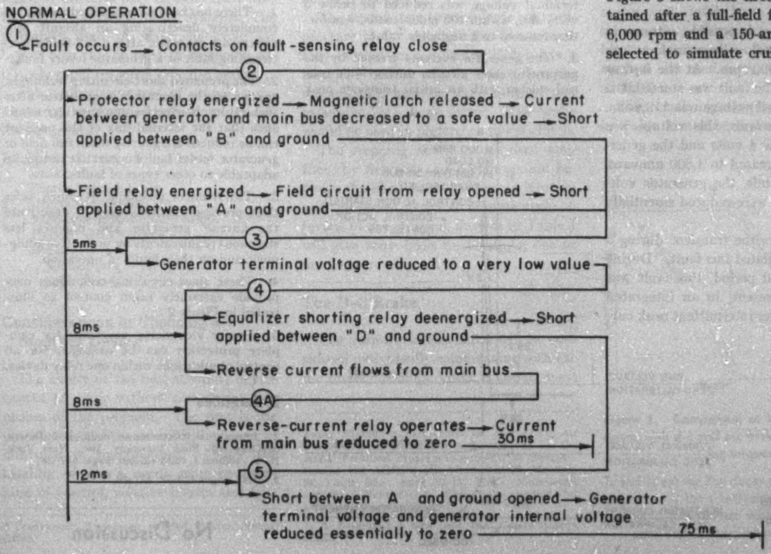

NORMAL OPERATION
①
Fault occurs → Contacts on fault-sensing relay close
②
Protector relay energized → Magnetic latch released → Current between generator and main bus decreased to a safe value → Short applied between "B" and ground
Field relay energized → Field circuit from relay opened → Short applied between "A" and ground
5ms
③
Generator terminal voltage reduced to a very low value
④
8ms
Equalizer shorting relay deenergized → Short applied between "D" and ground
Reverse current flows from main bus
8ms
4A
Reverse-current relay operates → Current from main bus reduced to zero
30ms
12ms
⑤
Short between A and ground opened → Generator terminal voltage and generator internal voltage reduced essentially to zero
75ms

Figure 2. Relay and circuit operational sequence from a fault

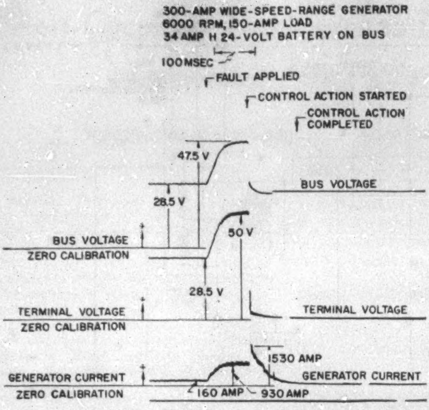


Figure 3. Full-field fault at simulated cruise conditions

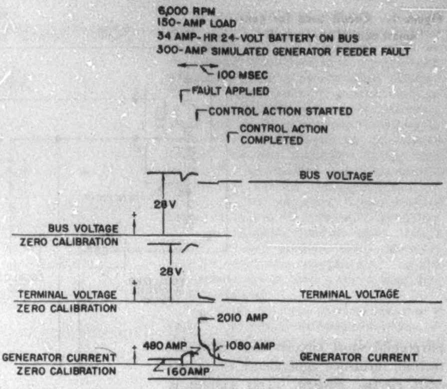


Figure 5. Simulated generator feeder fault

normal generator loading. At the instant that control of the fault was started, 930 amperes were flowing into the bus at 47.5 volts. Within 2 milliseconds, the high current between the generator and the bus was interrupted and the bus voltage was reduced to 28.5 volts. In that same time period, the generator voltage was reduced to 3 volts and the generator current increased to 1,530 amperes. Within 107 milliseconds, the generator current and terminal voltage were reduced essentially to zero.

Figure 4 shows a possible extreme in generator terminal voltage with a full-field fault at 9,000 rpm. At the instant that control of the fault was started, the generator terminal voltage was 113 volts. Within 2 milliseconds, this voltage was reduced to below 4 volts and the generator current increased to 1,800 amperes. In 100 milliseconds, the generator voltage and current were reduced essentially to zero.

Figure 5 shows the transient during a 300-ampere simulated bus fault. During its 43-millisecond period, this fault was intended to represent, in an integrated sense, much higher intermittent peak currents obtained during an arcing fault. At the instant that control of the fault started, the total generator current was 450 amperes. Within 2 milliseconds, the generator terminal voltage was reduced to less than 3 volts, and the generator current increased to 1,080 amperes. In 80 milliseconds, the generator voltage and current were reduced essentially to zero.

Similarities in these oscillograms and other oscillograms taken during the course of this investigation indicate that:

1. Within a few milliseconds after the fault-control action started, the generator terminal voltage was reduced to below 5 volts and, within 100 milliseconds, was further reduced to a negligible value.

2. The generator currents created by the purposeful short circuits during fault control started with an initial transient peak between 1,000 and 2,000 amperes and were reduced essentially to zero in 100 milliseconds.

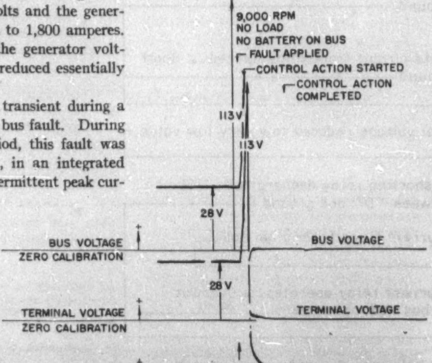


Figure 4. Full-field fault at no-load conditions

3. The bus-voltage transient after fault control started was negligible.

4. The generator current and voltage transients during the course of fault control showed relatively few differences after various types of faults and different conditions under which these faults were incurred.

Conclusions

1. There has been no recognized method of completely deactivating an aircraft self-excited d-c generator after a full-field short-circuiting fault or a generator feeder fault.

2. The presented short-circuiting technique, applied to the control of a generator after such faults, is effective and provides assurance that the internal flux of the machine can be reduced adequately when full-field or generator feeder faults occur. It is equally adaptable to other types of faults.

3. This control technique is much safer than previous methods used to deactivate the aircraft generator and imposes less strenuous requirements on protective equipment and on their limits of operation.

4. These short-circuiting techniques may provide extremely rapid control as illustrated in Figure 2.

5. Proper sequencing necessary for complete protection can be arranged for all control applications within one relay device.

References

1. Electrical Engineering (book), C. L. Dawes. McGraw-Hill Book Company, Inc., New York, N. Y., volume I, third edition, pages 437–38.

2. *Ibid*, pages 420–22, 431–36.

No Discussion

Stopping Devices for 400-Cycle Motors

L. A. ZAHORSKY
ASSOCIATE MEMBER AIEE

BECAUSE of its large size and complexity, the modern airplane requires many kinds of remotely or automatically controlled positioning systems.

Cooling air doors for engines, flight control surfaces, and engine speed control comprise but a few that require the electromechanical actuator to stop quickly and accurately. This requirement, although simple to state, apparently has been difficult to satisfy; for the number of devices used to attain it are very numerous and opinions as to their desirability vary widely. This requirement, stated in more detail, is that an actuating device must move its load to a certain place, stop it accurately at this place, and finally hold it there until called upon to move again.

In the majority of these positioning systems the stopping and holding device is incorporated at the highest speed portion of the prime mover since this leads to the least size of braking device. The same being true for motors, it follows that the stopping device usually is associated with the motor.

Although this paper is restricted to stopping devices for 400-cycle motors, much of the data may be associated with the similar problems for d-c systems. The principles used to make 400-cycle a-c brakes also can be used in designing extremely rapid response type d-c brakes.

The numerous types of stopping devices may be classified into one of the following groups:

1. Direct current
    - Brake
    - Fastop* clutch brake
2. Alternating current
    - Face-type motor clutch brake
    - Fastop* clutch brake
    - Brake
        - Shaded pole
        - Split phase

Considerations in Choosing a Stopping Device

The choice of the best stopping device cannot be made without considering two phases of the problem. The first is the user's needs and the second is the degree of difficulty to manufacture. Under the heading of user's needs may be listed kind of control, relative inertia and friction in the load, whether load is friction or overhauling, and availability of the device.

There are proportional, pulsed, and on-off controls. The proportional control uses a holding device only during periods of failure of power supply or control device. The pulsed systems, such as automatic temperature controls usually are best served by a clutch to allow the motor to stay continuously near its rated speed. The on-off controls may use any type of device depending on the following factors.

Friction load and inertia load are important factors in determining whether to use a Fastop clutch that stops the load only and allows the motor to coast, or a brake that stops the motor armature and load but has about three times the ability to hold the load. In general clutch brakes should be used with friction loads and brakes with overhauling loads. In some cases the availability of certain types of motor brake or motor Fastop clutch arrangements dictate the control. For example, the combination of a d-c brake with an a-c motor requires extra switching and power leads not needed on an a-c brake motor.

The second item to consider is the manufacture of the device. For example, it may be that an a-c Fastop clutch has a marginal benefit over a brake, but its manufacture may be more difficult and expensive by a large factor. The brake would be used in preference to the clutch.

The stopping device to be used may then, by necessity, be a compromise between the ideal device for the application and the easiest device to manufacture. A review of the stopping devices in Table I will give some basis to decide on one or more types.

The D-C Brake

The various devices listed will be considered individually commencing with the d-c brake. The d-c brake is the simplest of all the devices to manufacture but falls short of meeting the user's needs. It requires separate excitation from the motor and extra wiring and controls, as well as an inherent undesirable long flux-decay time caused by the eddy currents in its solid core.

During the period of flux decay, the motor is near its rated speed, and several revolutions of coast have occurred before the brake has started to resist the load. Figure 1 gives a comparison of speed versus time for a d-c and a-c brake of equal holding torque. This factor alone would limit the use of d-c brakes to a considerable degree.

The D-C Fastop Clutch Brake

The d-c Fastop clutch brake has been used on 400-cycle motors with considerable success. Its delay time is estimated about half that of the d-c brake due to its multiple air gaps. The fundamental difficulty is that with the separate excitation required by a d-c Fastop clutch, there is no change in current with load to help increase the slip torque of the clutch when loaded. When using a clutch on a d-c motor, a series coil may be incorporated to increase the torque capacity as the load increases, as shown by the diagram in Figure 2.

Since the current in the clutch must be sufficient to carry maximum torque at all loads, it is larger for a given torque rating than the d-c Fastop clutches. Since the motor and clutch have separate power sources, the user may have difficulty timing the clutch and motor starting. This method was used before a-c brakes were available.

The remaining stopping devices given in Table I may be connected to the a-c

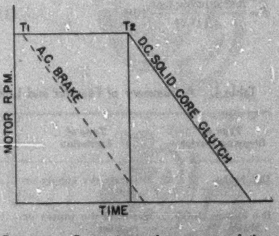


Figure 1. Comparison of motor speed time curves with d-c and a-c brakes of equal holding torques

T_1 and T_2 are the flux decay times before brake engages. T_1 for a laminated brake may be as low as 0.002 second whereas T_2 for solid core brakes approaches 0.05 second

---

* Trademark Registered United States Patent Office.

Paper 52-140, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE South West District Meeting, St. Louis, Mo., April 15-17, 1952. Manuscript submitted November 8, 1950; made available for printing February 27, 1952.

L. A. Zahorsky is with Lear, Inc., Grand Rapids, Mich.

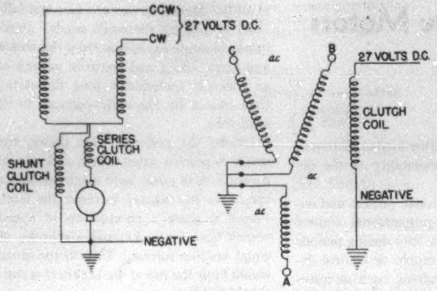


Figure 2. Connection diagrams of a-c and d-c Fastop clutches. Note that the current in the clutch is a function of load torque of the d-c motor

motor windings internally and require no additional external switching than the motor.

Obtaining Usable Pull from an A-C Device

Before discussing a-c brakes and clutches, the problem of obtaining a useable force must be defined. The two fundamental facts limiting the design of a-c braking devices are as follows:

1. The eddy-current losses in the solid iron core of a d-c brake are excessive at 400-cycle frequency and must be reduced to a workable value.
2. The pull of a single a-c flux path at 400 cycles is always positive and goes to zero 800 times per second. A quick glance at a general equation for pull will show this to be self-evident.

$$F=\frac{K\phi^2}{A} \tag{1}$$

ϕ for an a-c circuit is given by

$$\phi=\frac{e \text{ sine } 2\pi \text{ feet}}{4.44N}10^8 \tag{2}$$

Substituting equation 2 into 1 gives

$$F=\frac{Ke^2 \text{ sine}^2\, 2\pi \text{ feet}}{4.44^2N^2}10^{16} \tag{3}$$

Sine squared is always positive and goes to zero 800 times per second if $f=400$.

Overcoming Core Loss

The first of these two factors then requires that an efficient a-c brake or clutch uses a laminated structure with flux flowing parallel to the laminations at all times. Any flux crossing laminations of course would cause excessive excitation current and eddy-current losses. An attempt was made to develop a brake using a powdered iron coated with plastic to limit the eddy current losses.

The curves in Figures 3 and 4 show the results of this approach. At the operating force for a particular brake the ratio of watts for equal force was 3.91 with the brake closed. With the brake open the working force of 3 pounds was not attainable. Thus the magnetic path must be of high permeability in the direction of flux flow.

A somewhat similar set of curves was obtained with sintered iron which gives approximately twice the watts per pound pull. The necessary lamination of the core greatly limits the geometry of the a-c brake or clutch.

Methods of Obtaining Constant Pull

The second of these factors requires a phase shift of part of the air-gap flux to produce as nearly uniform pull as possible A-c motors and a-c brakes are very similar if motor torque and brake pull are thought of in a comparative sense.

What are the phase-shifting devices of a-c motors and their counterparts in the a-c brake? The same methods of phase shift may be employed by both except that some polyphase brake designs have been made using shaded poles whereas shaded-pole polyphase motors are practically nonexistent.

Among many of the phase-splitting devices used in single-phase motors are the shaded pole, capacitor split, capacitor start, and repulsion start induction run. The latter two types require rotation of the secondary to generate a speed voltage or to operate a centrifugal switch. A brake armature having no motion cannot utilize these two methods. Thus a-c motor brakes and clutches are left with shaded-pole or capacitor-split method only. Phase split by shading is done by introducing an I^2R loss in a portion of the flux path. Phase shift by capacitor is accomplished by cancelling inductive reactance with capacitive reactance.

From the standpoint of losses per unit of pull the capacitor-split brake has a fundamental advantage over the shaded-pole brake.

Various methods of obtaining low losses, constant pull, and a manufacturable device will be discussed later.

The Face-Type Motor Clutch Brake

The stator flux of the motor produces a large radial and constant component of force that is unused at present. Some axial force can be obtained by displacing the rotor, but the pull is nearly zero for considerable displacement and is zero

Table I. A Summary of Features and Limitations of Several Types of Stopping Devices Used in Conjunction with 400-Cycle A-C Motors

Type of Stopping Device	Type of Excitation	Kind of Special External Switching Required	Advantages over Other Types	Major Limitation or Defect
D-c brake	Separate d-c supply or rectifier	Extra contact on control relay	Simplest to fabricate	Brake does not respond as quickly as an a-c brake and requires separate d-c supply
D-c Fastop clutch brake	Separate d-c supply or rectifier	Separate remote control or extra contact on control relay	Combines clutch action with braking of load only	Holding ability low and require separate d-c supply
Face-type motor clutch brake	Motor stator winding is also brake winding	None	1. Brake stops load only 2. Only one electromagnetic structure 3. Minimum total losses of all a-c types	Motor must be short and large in diameter, requiring new actuator designs
A-c Fastop clutch brake	Motor stator and brake are connected internally	None	Brake stops load only	Fundamentally complex or uses brushes
Shaded-pole brake	Brake connected to motor leads internally	None	Requires no d-c power, faster response than d-c brake	Inherent high losses per unit torque. Complete and difficult construction
Split phase	Brake connected to motor internally	None	1. Lower inherent losses 2. Simpler to fabricate than shaded pole	Stops entire rotor

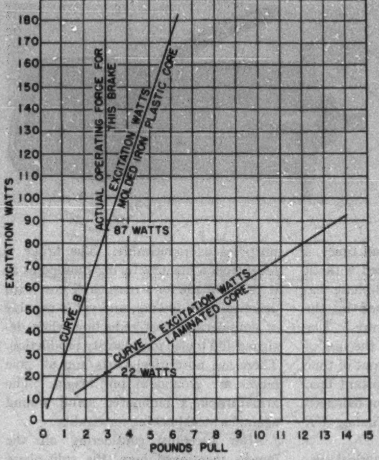


Figure 3 (left). Curves of excitation watts versus pounds pull in the normally closed position for laminated and plastic coated powdered iron core brakes of equal size

For 2-inch diameter 400-cycle brakes

Curve A—0.014-inch-thick laminated sheet steel

Curve B—molded plastic coated powdered iron

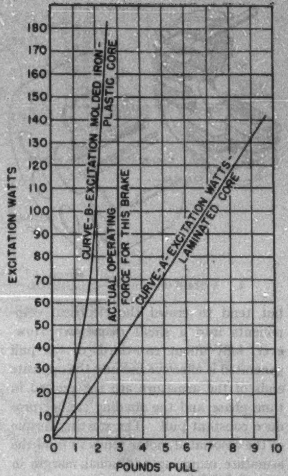
Figure 4 (right). Curves of excitation watts versus pounds pull in open position for laminated and plastic coated powdered iron core brakes of equal size

For 2-inch diameter 400-cycle brakes

Open position—0.015-inch air gap

Curve A—0.014-inch-thick laminated sheet steel

Curve B—molded plastic coated powdered iron



when the rotor is in its normal operating position. Also displacement reduces the effecting active length of the motor by twice the displacement.

A face-type motor would utilize all of the pull available in the air-gap flux. A preliminary study of such a motor immediately shows the possibility of very large forces at no excitation cost and without impairing the performance of the motor. As soon as the motor is energized the rotor moves to its designed running air gap to release the brake and engage the clutch. Earlier studies of face-type motors have indicated that this type of motor should have a large ratio of diameter to length compared to conventional drum-type motors. A diameter ratio in the order of 6 to 1 has been used successfully.

The actuator designs at present are predicated on motors whose diameter to length ratio is in the range of 0.2/1 to 0.4/1. Linear actuators usually have the motor in the same axis as the torque tube and screw. This arrangement presents a minimum of space required in one plane, thus allowing the actuator to be mounted in many close spaces. A face-type motor of large diameter mounted in this manner would increase the thinnest envelope dimension out of usable proportions. Mounting the motor at right angles to the gear-train axis would require a departure from conventional actuator design.

The face-type motor also would lend itself to very flat gearing for a rotary actuator, considering the large diameter in which to space the gearing, but requires basically new designs.

Therefore, the present configuration of both linear and rotary actuator must be abandoned to utilize the face-type motor.

The A-C Fastop Clutch Brake

The a-c Fastop clutch brake is the next item given that needs consideration at this time. The importance of the clutch to release the motor armature from the output shaft during braking was evidenced by the extensive use of the Fastop clutch on d-c motors. As a general rule the apparent inertia of a motored actuator is about 90 per cent motor and only 10 per cent load. Even in the instance of an aircraft gun turret, which is a very high inertia load, the actual measured data on one turret gave 94-per-cent motor inertia and 6-per-cent apparent load inertia. This accounts for the fact that Fastop clutches do not need to produce as much holding torque as a brake to accomplish the same stopping time for the usual type of loads.

At the present writing, all attempts to devise a Fastop clutch have led to either of two fundamental approaches. The first approach is to rotate the windings with the motor shaft and incorporate slip rings and brushes. The second is the use of a radially laminated rotating drive disk to carry the flux from the exciting poles to the driven disk. Figure 5 illustrates these two basic approaches. The desire to avoid the use of brushes on a-c motors and the problems of fabricating a radially laminated disk have prevented actual production of the a-c Fastop clutch.

The Shaded-Pole Brake

The brake shown in Figure 6 may be considered as three interrelated single-phase solenoids acting on two symmetrical armatures. Variations in pull are decreased by shading and by using alternate poles for return flux paths. The braking surfaces, being on the circumference of the brake, yield a maximum stopping effort for a given spring force

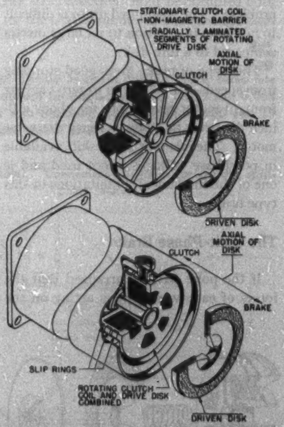


Figure 5. Two designs of a-c Fastop clutches. The disks are splined to the output pinion not shown

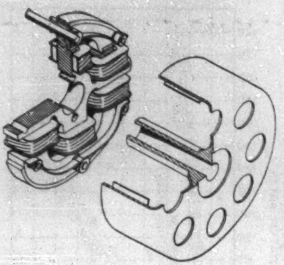


Figure 6 (left). Oblique sketch of a 3-phase shaded-pole brake

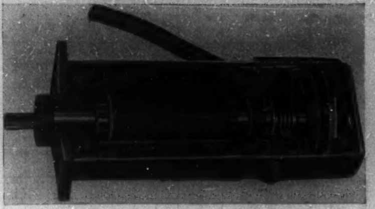


Figure 8 (right). Sectioned brake motor showing internal details of the split-phase brake applied to an actuator motor

but tend to crowd the electrical components into a small diameter. However, only about two-thirds of the pull generated is effective because the opposite ends of the armature are not excited in time phase and the shading cannot produce constant pull. The starting torque of the associated motor tends to bind the armature requiring additional margin in designed pull. The starting torque of the associated motor tends to bind the armature requiring additional margin in designed pull. The center pole must be ground shorter than the others by the sum of all manufacturing tolerances to insure the armature does not rock. This additional gap, plus the binding caused by starting torque and the inherent inefficiency of shading, all tend to increase the size and power consumption per unit of holding torque. The four carefully paired double-ended springs mounted on rollers holding the laminated armatures constitute a critical manufacture and assembly problem. The steel-lined drum is difficult to fabricate in addition to its large inertia compared to a thin disk.

This type of brake gives good stopping from speeds of 12,000 rpm to zero varying from 10 motor revolutions for about a 2-inch diameter to 30 for a 4-inch diameter motor. The inherent phase shift available in polyphase voltages is not used and is one of the fundamental limitations of this type brake.

The Split-Phase Brake

If the poles were so arranged that the forces of each were always acting on the spring and were balanced at all times, a constant pull could be obtained without resorting to shading with 3-phase excitation. This is accomplished by the split-phase brake shown in Figure 7. The total flux across the air gap is the sum of these equal fluxes 120 degrees out of time phase, resulting in exactly constant flux. This form of brake has constant balanced pull with no losses except winding I^2R and slight core loss.

The flux all travels in cylindrical surfaces both in the axis of the brake and circumferentially. Therefore, the core is laminated in essentially concentric rings. In 2-inch diameter motors and smaller, only four poles are used instead of the theoretically perfect six in order to get a space utilization of at least 90 per cent. With constant insulation thickness and clearances a 6-pole brake has better space utilization for motors larger than 2 inches in diameter.

The laminated core is integral with the steel-bearing insert and is pressed into the aluminum motor end bell to simplify construction and use all parts for as many purposes as practicable. The brake is ideal from a manufacturing standpoint as it can be made for a reasonable cost with practically no tooling, and yet is capable of considerable cost reduction when designed and tooled for quantity production. There are no critical parts and all pole pieces are ground at one height. The armature is a laminated scroll wound against a steel backing for rigidity.

After carefully considering all the factors mentioned here, the split-phase brake was adopted as the best compromise for immediate use on 400-cycle motors. Some performance and design data for a particular 4-pole 2-inch diameter brake are discussed to exemplify the possibilities of this type of brake.

Design

The simplicity of electrical design hardly merits discussion with the exception of the degree of saturation to use and the correct capacitor to obtain phase split. Although both items are capable of calculation it was found far more ex-

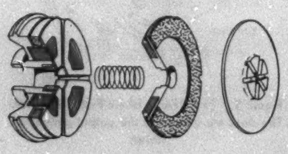


Figure 7 (left). Sectional drawing of a laminated split-phase 4-pole brake

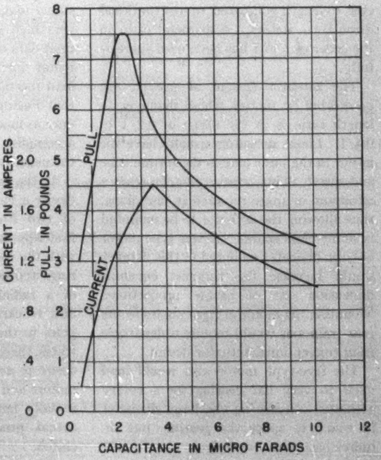


Figure 9 (right). Single-phase pull and current versus capacitance for a 2-inch diameter 4-pole brake

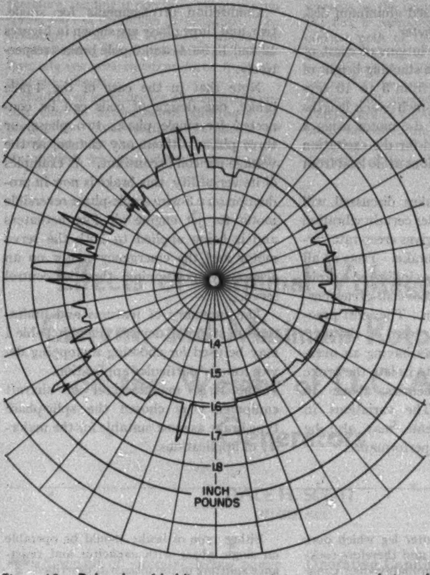


Figure 10. Polar plot of holding torque variation in one revolution of a typical brake

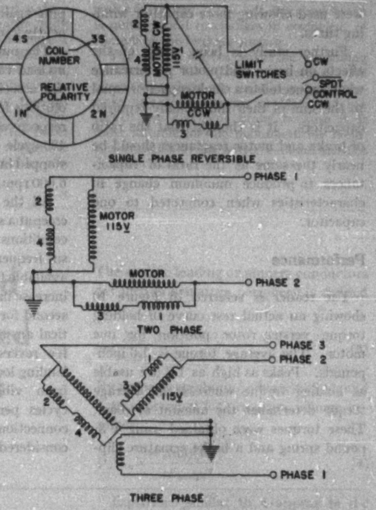

4S
3S
COIL NUMBER
RELATIVE POLARITY
1N
2N
2
4
MOTOR CW
115V
1
LIMIT SWITCHES
CW
SPDT CONTROL
CCW
MOTOR CCW
3
1
SINGLE PHASE REVERSIBLE
PHASE 1
2
MOTOR 115V
4
MOTOR
PHASE 2
3
1
TWO PHASE
PHASE 3
PHASE 2
1
3
115V
2
4
PHASE 1
THREE PHASE

Figure 12. Connection diagrams for 4-pole split-phase brakes

pedient to take a test curve to determine the best capacitor value.

Saturation should be based on a density suitable for operation with the worst combination, namely, two phases at 120 degrees. This allows the use of one set of coils for single- 2-, or 3-phase operation. The size of core was determined from the density at maximum voltage and minimum frequency whereas the spring tension is limited by the maximum spring tolerances associated with minimum voltage and maximum frequency. Therefore, the brake is designed to operate and test at 70 per cent of rated voltage.

When used on single-phase motors, a capacitor is required to obtain the phase split. A plot of brake pull versus capacitance is given in Figure 9. As might be expected a compromise capacitor of about 1 microfarad is used for this brake to prevent high-capacitor winding currents. A 3-lead reversible single-phase motor uses identical brake coils on the main and capacitor phase for simplicity of switching, whereas considerable capacitor size reduction could be obtained both for the brake and motor if reversing relays

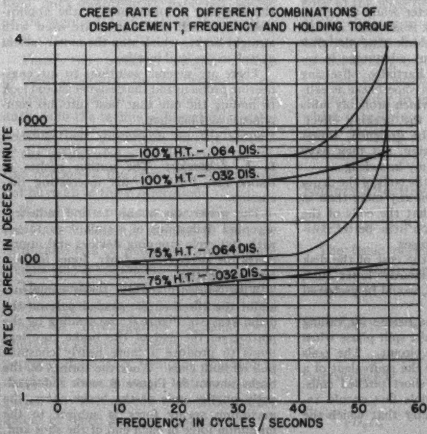
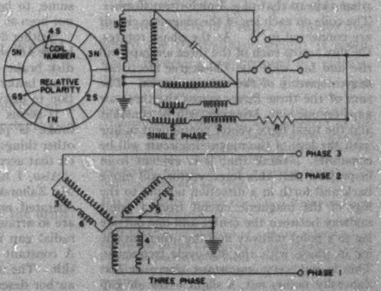


Figure 11 (left). Family of curves relating brake creep rate to vibration frequency for displacements of 0.032 and 0.064 total amplitude with opposing torque equal to 100 and 75 per cent of average static torque

Figure 13 (below). Connection diagrams for 6-pole split-phase brakes


4S
COIL NUMBER
5N
3N
RELATIVE POLARITY
6S
2S
1N
3
6
4
1
5
2
R
SINGLE PHASE
PHASE 3
PHASE 2
3
6
2
5
4
1
PHASE 1
THREE PHASE

were used allowing more capacitor winding turns.

Further study is being made of the effects on brake and motor performance when connected to a single capacitor equal to the sum of their individually required capacitors. It is thought that the ratio of brake and motor reactances should be nearly the same as the ratio of capacitances to produce minimum change in characteristics when connected to one capacitor.

Performance

The reader is referred to Figure 10 showing an actual test curve of holding torque versus rotor position for one motor. The average torque is 1.6 inch-pounds. Peaks as high as 1.8 are usable as holding torque whereas the average torque determines the amount of coast. These torques were obtained using a 3-pound spring and a brake armature slipping against an anodized aluminum disk nearly 2 inches in diameter.

The number of revolutions of coast at no load varies with the stacking height of the associated motor from 3 to 10 revolutions for 1/2- to 2-inch stack lengths respectively. A small d-c motor using a 400-cycle brake wound for d-c excitation stopped in $1^1/_2$ revolutions at no load from 6,000 rpm.

All the various brakes discussed will creep at a slow rate under certain vibration conditions. Figure 11 gives creep rate versus frequency for this brake. Tests on all available types of brakes indicated a rapid increase in creep rate above 40-cycles-per-second forced vibration. The most practical answer to this problem on the very few reversible actuators having an overhauling load has been to isolate the brake from vibration frequencies above 10 cycles per second. The variations in connection arrangements may also be considered as part of performance.

Connection arrangements for single, two, and three phase are shown in Figures 12 and 13 for 4- and 6-pole brakes respectively.

Note that in the case of the 4-pole brake, one design of coils can be connected for single phase, two phase, or three phase without any change, as the voltage per coil is constant. As examples of its versatility, this brake is now in production on a 3-wire single-phase reversible motor for jet engine air door actuators and is being applied to both the servo motor and the emergency motor on an accurately positioning throttle control servo.

The summary of features and limitations of stopping devices given in Table I may be used for choosing a stopping device for any particular application.

One large manufacturer of aircraft equipment has chosen the split-phase type brake as most suitable for the majority of applications.

Discussion

Ray D. Jones (General Electric Company, New York, N. Y.): Mr. Zahorsky is to be congratulated on a splendid paper on a timely subject of interest not only to the designers of aircraft electric motors but also to the users of those motors. As Mr. Zahorsky points out, brakes are a vital part of many aircraft electric motor applications.

On the question of different types of brakes, I do take some issue on Mr. Zahorsky's classification of the electric and magnetic circuits and also of the relative performance of the two types of 400-cycle actuated brakes.

I doubt if the terms "shaded pole" and "split phase" are proper appellations of these types when operated on three phase in the manner the author describes. The type he labels as shaded pole is not dependent, as I see it, on a time delay of some portion of the flux induced by shortcircuited turns around that portion of the flux.

Referring to Figure 6 of the paper, the magnetic circuit of this brake is somewhat akin to that of a 3-phase transformer. The coils on each leg of the magnetic circuit are connected in Y. As the phase rotates, the flux across each of the three air gaps will rise and fall with the 120-degree electrical-degree spacing of the voltage vector. The sum of the three fluxes across the three air gaps on one side, however, will be constant and the total force vector component in line with the legs of the magnetic circuit will be constant. I think that it is evident from inspection that this force vector will move back and forth in a direction normal to the legs of the magnetic circuit from a point midway between the center and one outside leg to a point midway and the other outside leg in phase with the 400-cycle frequency. This is not very important since, as Mr. Zahorsky points out, a slight extra air gap is introduced in the center leg which does not completely close up and therefore rocking cannot ensue. It probably does cause some noise.

I presume that Mr. Zahorsky's shaded section and shading coil are the dovetailed pole faces shown on the outside leg. This, however, shades the entire leg and I fail to see how this accomplishes a constant pull on that leg or that losses at this point add to the operation of the brake. In actual practice we have found improved operation, particularly in reduced loss on this type of brake, by eliminating as far as is practical eddy current losses at the pole faces of the brake.

While it is true that the drum has a large inertia as compared to the disk of the disk brake, the braking material is all working at the maximum diameter while the effective diameter of the disk is less than its outside diameter. I do think that the disk brake probably has some advantage in respect to the inertia. Partially offsetting this, however, the drum shoe brake is self-energizing to a degree which probably adds about 50 per cent to the braking effect. The disk brake could be designed, I presume, to be self-energizing, but they have not been so designed to my knowledge.

Neither do I see that the coils of the drum brake are squeezed down any more than in disk brake and think that the coils of the drum brake might be in a little better position to dissipate their losses.

It seems to me that the cost of the disk brake is probably inherently lower and, other things being equal, would be preferred on that account.

Also, I hardly see the reason for calling Mr. Zahorsky's disk brake split phase when operated on a 3-phase circuit. The coils are so arranged as to be the equivalent of a radial gap motor with short pitched coils. A constant rotating 2-pole flux should result. The action is exactly that which the author described.

Either type of brake should be operable on single phase with capacitor and resistance splitting of the phases as described.

In regard to the advantages of the clutch brake combination, I recognize, as Mr. Zahorsky points out, that the bulk of the inertia of a motored system is almost always in the motor rotor and I presume that for most applications some advantage in stopping time lies in the clutch brake. I would like to point out, however, that if the space occupied by the clutch brake were utilized in the straight brake, improved braking torque would ensue over the brake of the clutch brake alone which would compensate for some portion of the added inertia which this brake has to stop. We have found not too much difficulty in meeting the stopping requirements of the applications on which our motors are used with straight brakes which stop the motor inertia as well as the load inertia.

There are several solutions to an engineering problem and the designer should look to finding the one that best suits his components and his shop.

L. A. Zahorsky: The author wishes to thank Mr. Jones for his interesting discussion.

The writer was unable to find industry-accepted definitions of terminology to use with 400-cycle stopping devices and appreciates the suggestions of Mr. Jones for better classification.

With reference to Mr. Jones' comments about the effect of the shaded pole on the drum brake of Figure 6, the shading by the plating on the end of the brake core legs tends to produce a more nearly constant pull on both ends. Since the springs on the brake shown in Figure 6 work independently on the ends of the brake shoes, the maximum spring force is subject to the minimum force of each end of the shoe and, therefore, shading is necessary for maximum spring-force or maximum performance. The disk-type brake has symmetry about the point of spring force application and, therefore, always sees constant force from the magnetic circuit.

Shading of an entire leg of a magnetic circuit can be accomplished effectively by plating the entire surface. It should be noted that near the outside the flux is not surrounded by an eddy current path, whereas at the center of the plated surface the flux is surrounded by the entire eddy current path of the plating. Therefore a gradual shifting of phase of the flux will be obtained from the outside to the center, producing the same electrical effect as a shading coil around a portion of the leg. The feature of self-energization of the drum brake decreases the reliability of the brake releasing under low voltage and makes the operation of the brake more affected by corrosion in service. The self-energized brake is quite critical to changes in coefficient of friction of the braking surface.

The author agrees with the remaining portions of Mr. Jones' comments and agrees that there are usually several solutions to any engineering problem and has picked the disk-type brake because it is most suitable for the manufacturing facilities available to the writer without sacrifice of performance.

Effects of Terminal Voltage, Load Current, and Minimum Rotor Speed on the Weight of D-C Aircraft Generators

DALE H. SCOTT
MEMBER AIEE

THE weight of aircraft electric power systems is of major importance because unnecessary weight carried in an airplane reduces the pay load.

To make a weight study of aircraft electric power systems, it is necessary to obtain the weight of the electrical components. The generators which furnish the primary electric power are often the heaviest single units in the system.

Many factors affect the weight of d-c generators, among them environmental conditions encountered, type of cooling used, duty requirements expected, length of life required, and method of mounting used.

Thus, the terminal output ratings of these generators do not necessarily establish their weight. For this reason, the authors in the technical literature usually restrict their discussion on the subject of the weight of d-c generators to the active portion of the armature, which does not have all of these limitations.

For aircraft use, the operating conditions are sufficiently defined so that assumptions can be made and approximate weight curves obtained.

The methods used for obtaining the final weight curves are described. The range of investigation included was limited to existing generators so that experimental evidence could be obtained to check the assumptions. But the investigation can be extended to include higher values of speed and terminal voltage.

Operating Conditions

The operating requirements for aircraft generators are more severe than for more conventional types. Normally they are driven by the aircraft engine through a suitable gear box and thus must operate over a wide speed range as the engine speed is changed. This range is frequently a 3-to-1 speed change.

To reduce their weight, the generators are ventilated by forced convection. Openings are designed into the leading edge of the wing or in the forward surface of the airplane in such a manner that air can be admitted to the generator through special cooling tubes.

Since the smaller generators are attached rigidly to the gear box or gear pad which is built into the aircraft engine, extreme vibration is encountered, especially on reciprocating type of engines.

Factors Affecting Generator Weight

PERFORMANCE REQUIREMENTS

The basic performance equation of a d-c generator in terms of power developed in the rotor is

$$W_a = E_g I_a = K\phi n I_a = \frac{P\phi_p Z n I_a}{p(60)(10^8)} \qquad (1)$$

The average flux density over the armature surface is

$$B_{ave} = \frac{P\phi_p}{\pi D_a L_a} \qquad (2)$$

The specific loading or ampere conductors per inch of armature periphery is defined as

$$q = \frac{Z I_a}{p\pi D_a} \qquad (3)$$

By properly combining equations 1, 2, and 3 an expression for armature volume becomes

$$D_a^2 L_a = \frac{W_a(60)(10^8)}{n B_{ave} q \pi^2} \qquad (4)$$

Equations similar to 4 appear in d-c machine design textbooks.[1-3] Some authors have attempted to establish values of B_{ave} and q for typical generators.[2,3] From equation 4 it is clear that the power developed in the armature is in direct proportion to the armature volume.

TERMINAL VOLTAGE

In equation 4 no voltage term appears. Thus, for a given value of W_a it would appear that the voltage of the generator will not affect the volume of the rotor. As voltage increases, however, the insulation thickness must increase. Also, the number of conductors per inch of armature periphery will increase. Both of these factors will reduce the ratio of copper area to slot area, thereby decreasing the value of q for a given current density in the copper. As the voltage is increased, the rotor volume will increase in size for constant values of W_a, n, and B_{ave}.

As the rated voltage of the generator is increased, there is a range of values over which the commutator weight is essentially inversely proportional to voltage. The actual commutator diameter, the

---

Paper 52-142, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE South West District Meeting, St. Louis, Mo., April 15–17, 1952. Manuscript submitted August 8, 1951; made available for printing February 27, 1952.

DALE H. SCOTT is with the Naval Research Laboratory, Washington, D. C.

This paper is an abstract of a thesis submitted by the author to the University of Maryland Graduate School in partial fulfillment of the reqirements for the degree of Master of Science in Electrical Engineering.

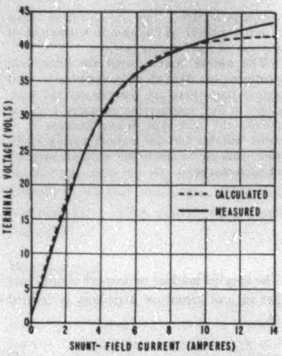


Figure 1. Comparison of measured and derived no-load saturation curves for basic generator

speed of rotation, the reactance voltages, and the volts per commutator bar are factors which impose limits on this inverse relation. If the diameter of the commutator exceeds a certain value, the centrifugal forces will be such as to require a more rigid design. As the voltage is increased, the increased armature weight will tend to compensate for a decreased commutator weight until structural limitations or reactance voltages impose a limit.

Type of Cooling

The type of cooling employed will determine the necessary surface area for ventilation. Aircraft generators are normally ventilated by forced convection. The restrictions in the ducts and generator will influence the amount of air which is used to cool the generator. Thus, the magnitude of B_{ave} and q in equation 4 are influenced by the rate at which heat is carried away.

Method of Mounting

The generators normally are flange-mounted on the engine drive pad in a cantilever fashion. Thus, the size of the yoke may be dictated by structural requirements rather than by flux-density requirements, and one end bell may also be heavier because of the necessity of supporting the entire generator weight while subject to vibration and to the gravitational changes encountered in aircraft operation.

Speed Range

The generators normally are operated over as much as a 3-to-1 speed range. The maximum speed of the rotor will determine certain structural requirements. The minimum speed will impose a limit on the magnetic circuit. The lower the required minimum rated speed, the greater will be the flux requirement to maintain rated terminal voltage and load current.

Method of Determining Weight Data

Design Study Proposed

The method chosen to obtain realistic weight data for aircraft generators was to design for specific ratings while retaining the flux densities, current densities, temperature rise of cooling air, and the operational speed range.

Basic Generator Selected

To establish the flux densities, current densities, speed range, and so forth, a specific aircraft generator was selected as a basic generator. The restrictions placed on the basic generator were that it should be a 30-volt 9-kw high-speed generator which was being used on naval aircraft. Within these limitations, the final choice of the basic generator was determined by availability. The determination of some of the constants resulted in the destruction of a unit. Thus, a generator type was selected of which identical units were available.

Design Method Explained

The design method used involved the determination of the constants of the basic generator. From the constants, the output of the basic generator was expressed in equation form. For each new design the values of flux densities, current densities, and temperature rise of cooling air were retained. The dimensional changes to retain the above values were then calculated, and the output equations determined for the designed generator. From the calculated dimensions, the change in volume for each component part was computed. The weight change for each part was then computed based on the mass density observed on the basic generator combined with the calculated volume change, or a standard mass density for the material combined with the calculated volume change, whichever was more logical to use. The changes in weight were then tabulated and the net weight change was combined with the weight of the basic generator to establish the weight of the designed generator. In this manner, only the changes were considered and a large amount of detail work was omitted.

Analysis of Basic Generator

Performance Equations

Predicting the performance of a d-c generator from design constants suggests that it should be possible to measure the constants of an existing generator and predict its performance. Errors in the method of prediction can then be detected by experimental test and the method corrected prior to using it to predict the performance of a generator which is being designed.

An equation for the no-load terminal voltage of a separately excited d-c generator was given by Van Valkenburg and Matthews[4] as

$$E_g = Kn[A \tanh(BN_fI_f+C)+D] \qquad (5)$$

By omitting the constant C, equation 5 becomes

$$E_g = Kn(A \tanh BN_fI_f+D) \qquad (6)$$

Evaluation of the constants should permit expressing the no-load saturation curve of the basic generator in mathematical form. The minimum speed of a blast-cooled aircraft d-c generator is defined as the lowest speed at which the generator will furnish rated load current at rated terminal voltage under stabilized temperature conditions with 1.5 ohms' resistance in series with the shunt field and under rated conditions of flow of cooling air.

Although the rated speed range of the generator was 4,550 to 8,000 rpm, the lower limit of speed or the minimum speed when measured was found to be 4,510 rpm. The accurate measurement of minimum speed was found to be extremely important in this study. The correct

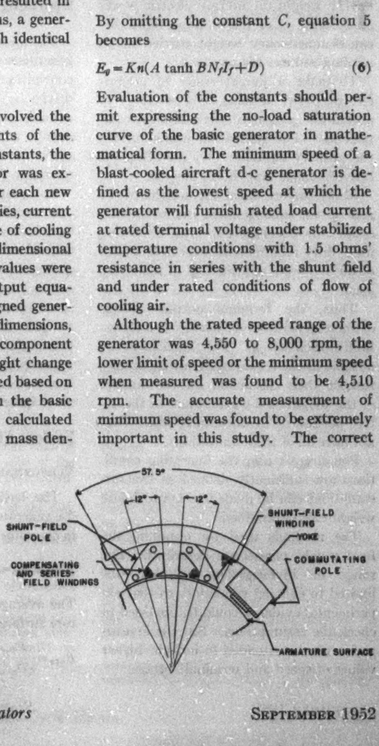


Figure 2. Partial end view of basic generator

values of air flow and series resistance must be used.

The measured curve of Figure 1 is the no-load saturation curve for the basic generator at 4,510 rpm. By substituting 40.5 for KnA, 1.2 for KnD, and 0.22 for BN_f, equation 6 becomes

$$E_g = 40.5 \tanh 0.22 I_f + 1.2 \qquad (7)$$

Equation 7 is the calculated curve of Figure 1. The shunt-field current, I_f, measured at full load and 4,510 rpm was 7.2 amperes. The calculated curve and the measured curve of Figure 1 agree reasonably well over the load range.

It has been demonstrated that the no-load saturation curve can be expressed in mathematical form. However, under conditions of load, equation 7 may not describe the relation between generated voltage and shunt-field current. The basic generator, as indicated in Figure 2, has compensating windings, series-field windings, and commutating poles.

Van Valkenburg and Matthews[4] also have developed an equation which expresses the generator terminal voltage in terms of direct-axis and quadrature-axis excitation and the armature circuit voltage drop. The equation modified only by omitting the same constant as was omitted from equation 5 to write equation 6 is

$$E_t = Kn\left\{\frac{A}{2BN_cI_a}\times \log_e\left[\frac{\cosh B(N_fI_f - N_dI_a + N_cI_a)}{\cosh B(N_fI_f - N_dI_a - N_cI_a)}\right] + D\right\} - I_aR_a \qquad (8)$$

The number of effective turns per pair of field poles furnishing quadrature-axis armature reaction is defined in equation form as

$$N_cI_a = N_c'I_a - N_c''I_a \qquad (9)$$

By observation N_c'' was determined to be 3.5 for the basic generator. From Figure 2, N_c' was also determined to be 3.5. Thus, full compensation is theoretically achieved. For either the case of armature current equal to zero or for perfect compensation, equation 8 becomes indeterminate. The limit of E_t, however, as I_a approaches zero is

$$\lim_{I_a \to 0} E_t = Kn(A \tanh BN_fI_f + D) = E_t \qquad (10)$$

Likewise

$$\lim_{N_c \to 0} E_t = Kn[A \tanh B(N_fI_f - N_dI_a) + D] - I_aR_a = E_t \qquad (11)$$

Since the basic generator is perfectly compensated, equation 11 becomes applicable. The direct-axis armature reaction N_dI_a is also made up of two terms and is defined by

$$N_dI_a = N_d'I_a + N_d''I_a \qquad (12)$$

From brush-shift measurement and series-field turns, for the basic generator

$$N_dI_a = 0.064 I_a \qquad (13)$$

Applying the measured values of N_f, R_a, and N_d for the basic generator, equation 11 becomes

$$E_t = 40.5 \tanh 0.001375\,(160\,I_f - 0.064 I_a) + 1.2 - 0.0231 I_a \qquad (14)$$

Figure 3 compares equation 14 with the measured full-load saturation curve and illustrates that it is possible to express the full-load terminal voltage of the generator in mathematical form if the no-load saturation curve and measurable constants are known. The equation is useful in adjusting the design constants.

PREDICTING EXCITATION

An analysis of the magnetic circuit of the basic generator was made to determine if the excitation required could be predicted from the physical dimensions of the generator and a knowledge of the physical properties of the materials. The results of the analysis showed that the required excitation at full load and minimum speed was 1,146 ampere turns per pair of field poles. The required excitation at full load and minimum speed calculated from equation 14 was 1,132 ampere turns per pair of field poles. The agreement is considered satisfactory and indicates that the method of checking the division of excitation is within reason.

TEMPERATURE RISE OF COOLING AIR

The total losses of the basic generator were determined by performing an efficiency test at full load and minimum rated speed while maintaining the proper mass rate of cooling air. The temperature rise of the cooling air was also measured.

COMPILATION OF DATA

Considerable additional data were obtained from the basic generator and were used for design purposes. A tabulation of the data was made and is included as Appendix III.

Discussion of Design

PARAMETERS RETAINED

Certain parameters were measured on the basic generator and retained for each design. These parameters were: current densities, average air-gap flux density, maximum rotor speed, minimum rotor speed, and temperature rise of cooling air. Retaining the current densities insured that the insulation used would be adequate from a thermal standpoint provided the heat was carried away at the same rate. Retaining the flux densities permitted the use of the same type of magnetic materials. Retaining the maximum rotor speed permitted the same type of structural design. Retaining the minimum speed and flux densities furnished a constant core loss per unit of volume. Retaining the temperature rise of cooling air insured essentially constant surface temperature over which the cooling air passed.

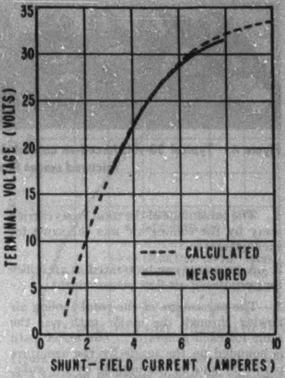


Figure 3. Comparison of measured and derived full-load saturation curves for basic generator

CURRENT DENSITY

The method used to retain current densities was to provide sufficient copper area so that the amperes per square inch of copper was the same as that measured on the basic generator.

AVERAGE AIR-GAP FLUX DENSITY

The average air-gap flux density was retained by determining the required generated voltage for the minimum-speed condition and providing the proper combination of conductors, armature surface, and excitation to retain the value of B_{avg}. An analysis was made of the excitation required for the armature, armature teeth, air gap, shunt-field poles, and the yoke in order to provide the proper excitation windings.

TEMPERATURE RISE OF COOLING AIR

To maintain the temperature rise of cooling air, the following assumptions were made:

Figure 4. Typical 30-volt forced-air cooled d-c aircraft generators; rated output of generators pictured ranges from 2.25 kw to 12 kw

1. The percentage of the total losses carried away by the cooling air was the same for each generator.

2. A generator can be treated as an orifice for purposes of air-flow measurement.

3. The percentage of the total cooling air flowing through the major path was the same for each generator. The major path is bounded on the inside by the armature periphery and on the outside by the yoke, the field poles, commutating poles, and windings.

Assumption 3 permitted defining the area of the major path for cooling air as A_v and was the only area calculated in order to account for all of the cooling air.

The assumptions permitted the following equation to be representative of all generators designed.

$$\frac{A_v}{W_a - W_o} = \text{constant} \qquad (15)$$

The constant of equation 15 from the basic generator data was 7.37×10^{-4} Thus, A_v for any generator becomes

$$A_v = 7.37 \times 10^{-4}(W_a - W_o) \qquad (16)$$

The dimensions of each generator designed were adjusted until equation 16 was satisfied.

Determining Shunt-Field Conductor Size

The method used to select the specific conductor size for the shunt-field winding is given in Appendix II. The method is quite general and permits an assumption for the value of excitation in order that the relations may be drawn. Changes in excitation requirements due to errors in assumptions may be corrected later by a fixed correction factor, which does not require redrawing the curves. The limitations on the power requirements of the voltage regulator are derivable from the curves.

Performance Equations

The no-load saturation curve can be plotted as soon as the shunt-field excitation is established and the generator constant K is known. The full-load saturation curve requires that the excitation furnished by the commutating winding, compensating winding, and series-field winding be established in addition to the no-load saturation curve; also, that the armature circuit resistance R_a be established. Then, the full-load saturation curve can be expressed in the form of equation 8 or equation 11, whichever is applicable.

Generator Weight

The physical dimensions of the generator have been established to write the performance equations. Therefore, based on the physical dimensions and the previously determined weights of the component parts of the basic generator, the change in weight of the component parts can be calculated. The results then can be tabulated and combined with the weight of the basic generator to give the new generator weight.

Design Results

Change in Voltage with Power and Speed Constant

The basic generator was rated at 30 volts, 9 kw, and had a minimum rated speed of 4,510 rpm. The effect of changing the voltage with constant output power and constant minimum speed was first obtained. Generators were designed for 60 volts and 120 volts. The weight of the 60-volt generator was identical to that of the 30-volt generator, but the 120-volt generator was approximately 2 pounds heavier. For the 60-volt generator there was a reduction in commutator weight, but the poorer space factor in the armature caused increased armature volume. This, in turn, increased the over-all generator diameter. The increased volume caused a weight increase which compensated for the weight reduction of the commutator. As a result the over-all weight did not change. For the 120-volt generator, the commutator weight reduction was not achieved because of the increased number of commutator bars.

Change in Power with Voltage and Speed Constant

A 30-volt 12-kw generator was designed. Its weight for the base speed of 4,510 rpm was 59.5 pounds. It should be kept in mind that all of the generators designed had the same speed range. This means the same maximum speed, which requires greater mechanical strength and, therefore, greater weight than would be necessary if only the base speed were considered.

Two additional generators were designed, one at 120 volts, 20 kw, and 4,510 rpm; the other at 120 volts, 30 kw, and 4,510 rpm. Their weights were 82 pounds and 106 pounds respectively. The tabulation of the 30-kw design is given in Appendix IV.

Experimental Verification of Design Results

Method

The most certain method of checking a design is to have the unit built to the design specifications and then test to determine conformance with predicted performance. Such a procedure was not practical in this case. Therefore, a substitute method was devised. A number of 30-volt generators were available, see Figure 4, but to conform with design data, only those having a maximum speed rating of 8,000 rpm were selected.

The base or minimum speeds of 12 generators were measured to furnish a common basis of comparison. Some of the generators furnished rated power output at speeds considerably below their nominally rated minimum; others would not furnish rated power at their nominally rated minimum speeds. Thus, a uniform method of obtaining the minimum speed was used which eliminated some variation in manufacturing and design.

From the minimum speed tests the

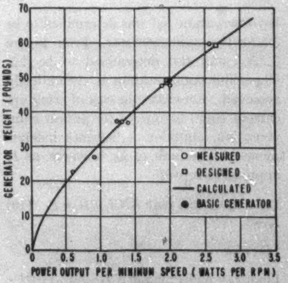


Figure 5. Generator weight as a function of power output per minimum speed. Comparison of experimental results and derived results for 30-volt generators

watts output per minimum revolutions per minute, W_o/n_1, was computed. The total weight of each generator was then obtained. The plot of generator weight W as a function of W_o/n_1 is shown in Figure 5.

ANALYSIS OF EXPERIMENTAL WEIGHT RELATIONS

The experimental data of Figure 5 are very close to linear. Equation 17 is written as an approximation of the relation in linear form.

$$W = 18.7(W_o/n_1) + 12 \qquad (17)$$

where W is the total generator weight expressed in pounds. The linear relation is not valid, however, for low values of W_o/n_1, for it is reasonable to expect that some power output should be available at less than 12 pounds. It is also reasonable to expect that as the value of W_o/n_1 is increased, the ratio of weight to W_o/n_1 should decrease, rather than remain a constant as equation 17 would require. Therefore, a curve is suggested in which W is proportional to some fractional power of W_o/n_1.

Derivation Involving Design Results

WEIGHT CURVE FOR 30-VOLT GENERATORS—POWER OUTPUT AND MINIMUM SPEED VARIED

From design data, the weight of the 12-kw 30-volt generator was 59.5 pounds, and the value of W_o/n_1 was 2.66. Comparable values of W and W_o/n_1 were obtained from Appendix III for the basic generator. From these two points and the origin, a curve was derived and is given as

$$W = 31(W_o/n_1)^{0.67} \qquad (18)$$

Equation 18 is included on Figure 5 and is more realistic than the linear relation of equation 17 at small and large values of output per minimum speed. It is generally conceded that 400 amperes is a practical upper limit of current which can be commutated successfully with no change in basic design. Therefore, 12 kw is an upper limit of power for which equation 18 will be accurate.

It is recognized that equation 18 was derived from design values in which only the load current was varied. The terminal voltage was 30 volts and the minimum speed was 4,510 rpm in each case. However, the experimental results of Figure 5 were obtained over a range of minimum speeds from 3,300 to 4,995 rpm. The agreement between equation 18 and the experimental values of Figure 5 is good. It is concluded, therefore, that equation 18 will be valid for values of minimum speed varying between 3,000 and 8,000 rpm. The upper limit of 8,000 rpm is due to structural limitations.

WEIGHT CURVE FOR 120-VOLT GENERATORS—POWER OUTPUT AND MINIMUM SPEED VARIED

The weights of the 9-kw, 20-kw and 30-kw generators were 51, 82, and 106 pounds respectively. The values of W_o/n_1 were 2.00, 4.43, and 6.65 watts output per minimum revolutions per minute respectively. The preceding points and the origin were used to derive an equation in the same manner as for the 30-volt generators. The resulting equation is

$$W = 33.4(W_o/n_1)^{0.61} \qquad (19)$$

Equations 18 and 19 are shown graphically in Figure 6 for comparison.

As in the 30-volt designs, the minimum speed was 4,510 rpm for all of the 120-volt generators designed. The range of minimum speeds expected is the same as for the 30-volt generators. It was shown experimentally for the 30-volt generators that the effect of varying the speed was also included in equation 18. Speed appears in equation 19 as a variable.

VERIFICATION OF 120-VOLT WEIGHT CURVE

No generators were available to check equation 19 experimentally. However, the weight and minimum speed were obtained for one 30-kw generator, which was tested by another agency. The weight was 140 pounds and the minimum speed was 2,900 rpm. Thus, W_o/n_1 was 10.33. From Figure 6, the weight of a 30-kw 120-volt generator whose watts output per minimum revolutions per minute is 10.33 should be between 138 and 139 pounds. Therefore, in this particular case, the variation between the predicted weight and the actual weight was less than 2 pounds, which is less than 1.5 per cent. The same method of design was used for both the 30-volt and 120-volt generators. The experimental verification of the 30-volt curve established the method. Therefore, less experimental evidence is required to prove the validity of the 120-volt curve.

Conclusions

It has been demonstrated that the effects of changes in terminal voltage, minimum rotor speed, and load current on the weight of d-c aircraft generators can be expressed in mathematical form. Equations 18 and 19 express these mathematical relations, which are also shown graphically in Figure 6. The following limitations should be noted in the use of the curves of Figure 6:

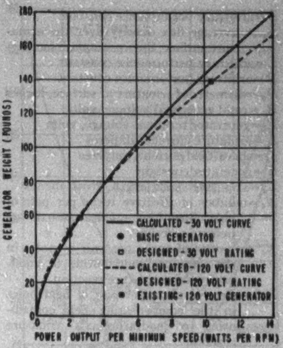


Figure 6. Derived results of generator weight as a function of power output per minimum speed for 30-volt and 120-volt generators

1. An upper limit of 400 amperes load current should not be exceeded.
2. An upper limit of 8,000 rpm is the maximum rated speed; therefore, the minimum speed has an upper limit of 8,000 rpm. A correction factor must be applied to the curves if it is desired to know the weight of generators whose maximum speed is greater than 8,000 rpm.
3. No definite lower limit has been set on the minimum speed; in general, however, the interest is in minimum speeds in excess of 3,000 rpm.

For any fixed power output, the effect of minimum rotor speed can be determined by reading the values from the proper curve of Figure 6.

For low ratings of power output per minimum speed, the 30-volt generators will weigh less than the 120-volt units. For higher ratings of power output per minimum speed, a weight saving will result from the use of 120-volt generators.

The effect of varying only the terminal voltage is available from Figure 6 by interpolating between the limits of 30 and 120 volts.

The relations of Figure 6 are approximate, and are limited to the type of design used in the basic generator, but it is believed that they are sufficiently representative to furnish practical values for use in aircraft electric system weight studies. Such values have not heretofore been available.

Appendix I. List of Symbols

A = arbitrary performance constant
A_v = ventilation area, square inches

B = arbitrary performance constant
B_{ave} = average flux density over the armature surface, lines per square inch
C = arbitrary performance constant
D = arbitrary performance constant
D_a = diameter of armature at surface, inches
E_g = total generated voltage, volts
E_t = generator terminal-voltage, volts
I_a = armature current, amperes
I_f = shunt-field current, amperes
K = generator constant
L_a = armature-stack length, gross, inches
N_c = number of effective turns per pair of poles furnishing quadrature-axis armature reaction
N_c' = number of turns per pair of poles in the armature contributing to quadrature-axis armature reaction
N_c'' = effective turns per pair of field poles in the compensating field contributing to quadrature-axis armature reaction
N_d = number of effective turns per pair of poles furnishing direct-axis armature reaction
N_d' = number of turns per pair of poles in the armature contributing to direct-axis armature reaction
N_d'' = effective series-field turns per pair of poles contributing to direct-axis armature reaction
N_f = number of turns in shunt field per pair of poles
n = speed of generator rotor, revolutions per minute
n_1 = base speed or minimum rated speed of generator rotor, revolutions per minute
P = number of shunt-field poles
p = number of current paths in generator armature
q = ampere conductors per inch of armature periphery
R_a = total resistance of armature circuit including brush drop and resistances of series field, commutating windings, and compensating windings, ohms
R_f = resistance of shunt field, ohms
W = generator weight, pounds
W_a = power developed in armature rotor, watts
W_o = power output of generator, watts
Z = number of armature conductors
ϕ = total armature surface flux, lines
ϕ_p = total armature surface flux divided by the number of poles, lines

Appendix II. Determining Shunt-Field Conductor Size

Figure 7 is a graph for selecting the conductor size to use on a generator once the value of N_fI_f is established. The method used to draw the curves is given as follows:

1. The current density was assumed to be the same as for the basic generator. Thus, from standard tables giving dimensions of wire sizes the current curve was plotted.

2. The value of N_fI_f was assumed constant at 1,152 ampere-turns. Thus, from the ampere-turns assumption and equation 1, the turns curve was plotted.

3. A standard space-factor table for enameled magnet wire was used. From the standard value of space factor and equation 2, the area per pair of poles was calculated and drawn.

4. The value of $I_f^2R_f$ was measured for the basic generator and was assumed constant. From the value of $I_f^2R_f$ and equation 1, R_f and I_fR_f were calculated and plotted.

A set of curves was then available for determining the shunt-field conductor size.

In the event the excitation requirements increase after drawing the curves, the new values of current and turns should be established. Then the new wire size can be selected. The new values of resistance, voltage, power, and area can be obtained by reading the number from Figure 7 for the wire size selected and multiplying by the ratio of the new to the old turns.

Appendix III. Tabulation of Basic Generator Data

Description	Symbol	Explanation or Numerical Value	Units
Armature diameter	D_a	3.919	inches
Armature stack length	L_a	3.91	inches
Number of armature slots	S	37	
Number of armature conductors	Z	74	
Armature conductor area	C_{aa}	0.023	square inches
Number of current paths through armature	p	2	
Armature current at base speed	I_a	307.2	amperes
Armature conductor width	C_{wa}	0.095	inches
Armature conductor depth	C_{da}	0.242	inches
Armature slot width	s	0.115	inches
Armature average tooth width	t	0.156	inches
Average flux per pole at base speed	ϕ_p	343,000	lines
Number of shunt-field poles	P	4	
Axial length of shunt-field pole	L_f	3.84	inches
Clearance between armature and shunt-field-pole shoe	δ	0.0115	inches
Inside diameter of yoke	D_{yi}	5.328	inches
Outside diameter of yoke	D_{yo}	6.00	inches
Axial length of yoke	L_y	6.25	inches
Number of commutating poles		4	
Type of compensating winding		concentrated	
Commutator diameter	D_c	2.625	inches
Axial length of commutator	L_c	2.468	inches
Number of commutator bars	Z_c	37	
Volts per bar, approximate		4.3	volts
Total number of brushes		8	
Brush thickness		0.4375	inches
Brush width		0.9375	inches
Brush current density		187	amperes per square inch
Current per brush at base speed	I_b	76.8	amperes
Shunt-field current at base speed	I_f	7.2	amperes
Shunt-field resistance at base-speed temperature	R_f	2.95	ohms
Shunt-field turns per pair of poles	N_f	160	turns
Armature resistance at base-speed temperature	R_{aa}	0.00733	ohms
Series-field plus compensating-winding resistance at base-speed temperature	R_{ac}	0.00619	ohms
Commutating-winding resistance at base-speed temperature	R_{at}	0.00303	ohms
Total armature circuit resistance at base-speed temperature $= R_{aa}+R_{ac}+R_{at}$	R_{a1}	0.0166	ohms
Efficiency		76	percent
Power output per revolutions per minute at base speed	W_o/n_1	1.995	watts per revolutions per minute
Total generator weight	W	49.2	pounds
Base speed	n_1	4510	revolutions per minute

Appendix IV. Tabulation of Data for 120-Volt 30-Kw Generator

Description	Symbol	Explanation or Numerical Value	Units
Armature diameter	D_a	6.50	inches
Armature stack length	L_a	4.07	inches
Number of armature slots	S	75	
Number of armature conductors	Z	150	
Armature conductor area	C_{aa}	0.019	square inches

Number of current paths through armature	p	2	
Armature current at base speed	I_a	255.5	amperes
Armature conductor width	C_{wa}	0.0872	inches
Armature conductor depth	C_{da}	0.218	inches
Armature slot width	s	0.107	inches
Armature average tooth width	t	0.136	inches
Average flux per pole at base speed	ϕ_p	395,000	lines
Number of shunt-field poles	P	6	
Axial length of shunt-field pole	L_f	4.00	inches
Clearance distance between armature and shunt-field-pole shoe	δ	0.040	inches
Inside diameter of yoke	D_{yi}	8.52	inches
Outside diameter of yoke	D_{yo}	9.28	inches
Axial length of yoke	L_y	6.41	inches
Number of commutating poles		6	
Type of compensating winding		concentrated	
Commutator diameter	D_c	4.35	inches
Axial length of commutator	L_c	1.66	inches
Number of commutator bars	Z_c	75	
Volts per bar, approximate		12	volts
Total number of brushes		12	
Brush thickness		0.341	inches
Brush width		0.669	inches
Brush current density		187	amperes per square inch
Current per brush at base speed	I_b	42.6	amperes
Shunt-field current at base speed	I_f	5.5	amperes
Shunt-field resistance at base-speed temperature	R_f	17.2	ohms
Shunt-field turns per pair of poles	N_f	452	turns
Armature resistance at base-speed temperature	R_{aa}	0.0190	ohms
Series-field plus compensating-winding resistance at base-speed temperature	R_{ac}	0.0218	ohms
Commutating-winding resistance at base-speed temperature	R_{at}	0.0069	ohms
Total armature circuit resistance at base-speed temperature $= R_{aa}+R_{ac}+R_{at}$	R_{a1}	0.0475	ohms
Efficiency		85	per cent
Power output per revolution per minute at base speed	W_o/n_1	6.65	watts per revolutions per minute
Total generator weight	W	106	pounds
Base speed	n_1	4510	revolutions per minute

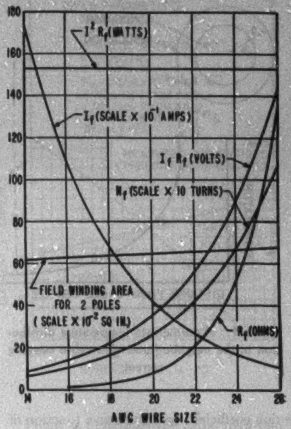


Figure 7. Graph for determining shunt-field conductor size

References

1. Elements of Electrical Design (book), Alfred Still. McGraw-Hill Book Company, Inc., New York, N. Y., 2nd edition, 1932.

2. Design of Electrical Apparatus (book), John H. Kuhlmann. John Wiley and Sons, Inc., New York, N. Y., 2nd edition, 1940.

3. Aircraft Electrical Engineering (book), F. G. Spreadbury. Sir Isaac Pitman and Sons, Limited, London, England, 1943.

4. Steady-State Analysis of Aircraft D-C Generators, Ernest Van Valkenburg, Whitney Matthews. *AIEE Transactions*, volume 67, part II, 1948, pages 1351–58.

No Discussion

Continuous Current and Temperature Rise in Aircraft Cables

M. SCHACH
ASSOCIATE MEMBER AIEE

THE current capacity of aircraft cables in the simplest of environmental conditions is governed by many variables whose relations cannot be made explicit easily. For this reason, most current ratings have been determined empirically. The desirability of having a method for calculating current ratings becomes evident when changes are made in cable or in the cable environment. Improved insulation, lighter conductors, and higher ambient and operating temperatures are some of the more important factors which affect current ratings.

This paper describes two methods by which the continuous current capacity of a single cable in air may be predicted when the limiting condition is a maximum conductor temperature. In the first, the current is found by the simultaneous graphical solution of the equations for the transmission of heat (1) through the insulation and (2) from the outer cable surface. In the second method, the two heat transfer relations mentioned previously are combined into a single equation which may be used to determine the outer cable surface temperature. When this temperature is known the current can be calculated with either of two heat transfer relations applicable to the cable.

Both methods require that the convec-

Paper 52-141, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE South West District Meeting, St. Louis, Mo., April 15–17, 1952. Manuscript submitted February 19, 1951; made available for printing February 27, 1952.

M. Schach is with the Naval Research Laboratory, Washington, D. C.

The author wishes to acknowledge the aid of W. K. Gardner and R. Kidwell of the Naval Research Laboratory for conducting the experimental work required to obtain the convection coefficients.

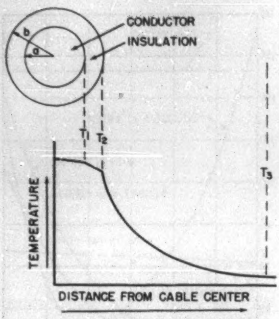


Figure 1. Approximate temperature distribution in the neighborhood of a cable carrying current

tion coefficient be known as a function of (1) the cable diameter and (2) the temperature difference between the outer cable surface and ambient air. The convection coefficients were determined experimentally for the range of diameter and temperature difference normally encountered. Measurements were obtained at air pressures corresponding to sea level, 20,000, and 60,000 feet.

At the same time that the convection coefficients were obtained the effect of altitude on current capacity was measured for typical aircraft cables. Results of these measurements will be given.

It will be shown that the second method for determining current also provides a convenient means of estimating the effect of changes in any of the cable parameters on the current capacity of the cable.

Assumptions and Review of Heat Transfer Relations

When an electric current flows in a conductor the heat generated is transmitted through the insulation by conduction and then transferred to the surrounding air and walls by convection and radiation. If the cable insulation is to be prevented from deteriorating too rapidly and safe operation of electric circuits is to be maintained, the insulation temperature must not exceed some predetermined value. Because the conductor temperature is very nearly uniform, and the temperature of the insulation approaches the conductor temperature at their common boundary, the thermal limitation on the current capacity can be stated as a maximum conductor temperature. Maximum conductor temperature is the only criterion affecting current capacity which will be considered in this analysis.

The following assumptions will be made:

1. The cable is suspended in air.
2. It consists of a cylindrical conductor with one or more concentric layers of insulation.
3. Steady-state conditions apply.
4. Heat transfer from the outer cable surface is by natural convection and radiation.

The rate at which heat is produced in the conductor and conducted radially through the insulation per unit length of cable is given by

$$P=I^2R_{T_1}=\frac{2\pi K(T_1-T_2)}{\log_e(b/a)} \tag{1}$$

where

I = current
R_{T_1} = conductor resistance per unit length at temperature T_1
K = thermal conductivity of the insulation
T_1 = conductor temperature
T_2 = temperature of the outer cable surface
b = over-all cable radius
a = conductor radius

The rate at which heat is transferred from the outer cable surface by convection per unit length can be expressed as

$$P_c=2\pi bh(T_2-T_3) \tag{2}$$

where

h = convection coefficient
T_3 = temperature of the air at a point where the temperature gradient has become negligibly small. Figure 1 indicates the approximate temperature distribution in the neighborhood of a cable carrying current

Finally, the power radiated per unit length from the cable exterior may be expressed by

$$P_r=2\pi bse(T_{2k}{}^4-T_{3k}{}^4) \tag{3}$$

where

s = Stefan's constant
e = emissivity
T_{2k} = temperature of the outer insulation surface in degrees absolute
T_{3k} = temperature of the surrounding walls in degrees absolute. It will be assumed that the wall temperature is the same as that of the air at a point where the temperature gradient has become negligible.

If the radiation coefficient r is defined to be the power radiated by unit area for unit difference in temperature between the cable surface and the ambient air, that is

$$r=\frac{se(T_{2k}{}^4-T_{3k}{}^4)}{T_2-T_3} \tag{4}$$

the power radiated per unit length can be rewritten as

$$P_r=2\pi br(T_2-T_3) \tag{5}$$

The total power transferred per unit length from the cable exterior becomes

$$P=P_c+P_r=I^2R_{T_1}=2\pi b(h+r)(T_2-T_3) \tag{6}$$

Predicting the Current

We now wish to determine the current which will produce a given conductor temperature T_1 in a particular cable suspended in air at a specified ambient temperature T_3. It will be assumed for simplicity that the insulation consists of a single layer. One expression for the current may be obtained from equation 1.

$$I=\left[\frac{2\pi K(T_1-T_2)}{R_{T_1}\log_e(b/a)}\right]^{1/2} \tag{7}$$

In this equation T_1, R_{T_1}, b, and a are given constants while K is a function of insulation temperature.

Because experimental measurements show K changes slowly with temperature and small changes in its value affect the current capacity in a negligible way, K will be considered as a constant for any given insulation. The effect of a change in K on the current rating will be illustrated later. The first step in finding the

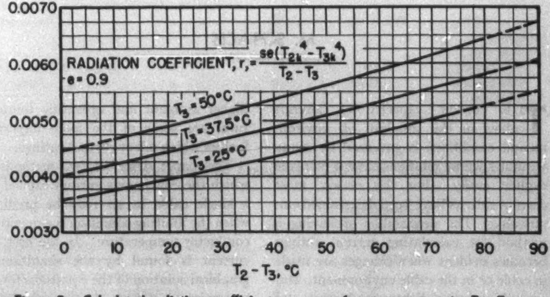


Figure 2. Calculated radiation coefficient r versus surface temperature rise T_2—T_3

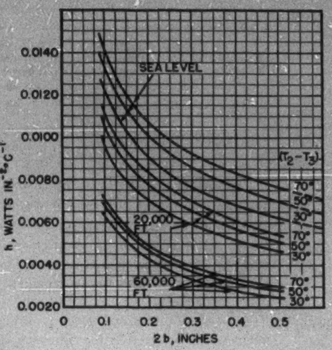
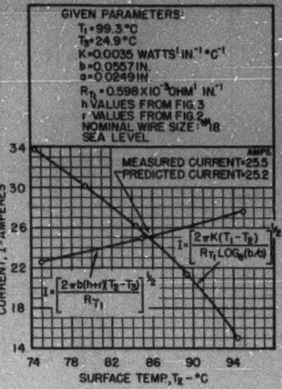


Figure 3 (left). Measured convection coefficient h versus over-all cable diameter 2b, for sea level and simulated altitude conditions

Figure 4 (right). Graphical solution for the current satisfying the given conditions of heat transfer. Nominal wire size: AN 18



current is to plot I as a function of T_2 with the aid of equation 7.

A second expression for the current can be obtained from equation 6.

$$I=\left[\frac{2\pi b(h+r)(T_2-T_3)}{R_{T_1}}\right]^{1/2} \qquad (8)$$

In this equation b, T_3, and R_{T_1} are given constants; r is a function of (T_2-T_3) and T_3; h is a function of T_2-T_3 and cable diameter $2b$ for any specified air density.

For each value of T_2 arbitrarily selected we find the corresponding value of r from equation 4. To apply equation 4 the emissivity of the outer cable surface must be known. Since the emissivity of most nonmetallic materials lies between 0.8 and 1.0, e may be assumed to have the value 0.9 without introducing a large error. The calculated r values for the range needed are plotted in Figure 2. The convection coefficient h was determined experimentally as a function of (T_2-T_3) and b. A description of the method by which h was determined is given in Appendix I.

Figure 3 is a plot of h as a function of $2b$ for various (T_2-T_3) values at pressures corresponding to sea level, 20,000, and 60,000 feet. For the curves shown, the convection loss was found by subtracting the computed radiation loss from the measured total dissipation. Figure 3 is used to find the value of h corresponding to an arbitrarily chosen T_2. The current now can be plotted as a function of T_2 with the aid of equation 8. The intersection of the two curves of current versus T_2 determines the desired current since it satisfies the two conditions of heat transfer.

The method described has been applied to size AN 18, AN 8, and AN 0 aircraft cables in Figures 4, 5, and 6 respectively. Because values of K were not available they had to be obtained by measuring b, a, P, T_1, and T_3 and substituting in equation 1. The experimental method used to determine K is described in Appendix I. Since the insulation of each cable consisted of a layer of plastic and a layer of impregnated cotton braid, the thermal conductivity found was an effective value for the composite insulation. Values for r and h were taken from Figures 2 and 3. The computation details for the case illustrated in Figure 4 are given in Appendix II. The actual measured current shown in each figure demonstrates the degree of accuracy obtained by using the graphical solution. Methods which make use of the same principle have been applied in determining power cable ratings and to the problem of computing the heat loss from insulated pipes.[1–3]

The method of finding the current just described may be extended to the case where the insulation consists of two or more layers of insulation. If there are two layers with appreciably different thermal conductivities, three expressions for the current may be obtained. For this case

$$I=\left[\frac{2\pi K_1(T_1-T_2')}{R_{T_1}\log_e(b'/a)}\right]^{1/2} \qquad (9)$$

$$I=\left[\frac{2\pi K_2(T_2'-T_2)}{R_{T_1}\log_e(b/b')}\right]^{1/2} \qquad (10)$$

$$I=\left[\frac{2\pi b}{R_{T_1}}(h+r)(T_2-T_3)\right]^{1/2} \qquad (11)$$

where equations 9 and 10 come from the heat conduction relations for the inner and outer insulation layers respectively. The new symbols in these equations are:

K_1 = thermal conductivity of the inner layer
K_2 = thermal conductivity of the outer layer
T_2' = temperature at the boundary between the two layers
b' = radius from the cable center to the boundary between the two layers

Figure 7 illustrates the cable with two layers of insulation. Equation 11 is identical with equation 8. To find the current which will produce a given T_1, equations 9 and 11 are used to plot I as a function of T_2' and T_2 respectively. From these two curves one pair of T_2' and T_2 values is obtained for each current value. If these paired values of T_2' and T_2 now are substituted in equation 10, and I is plotted as a function of T_2, a curve is obtained whose intersection with that of equation 11 gives the desired current.

For the general case of multilayer insulation, there will be as many expressions

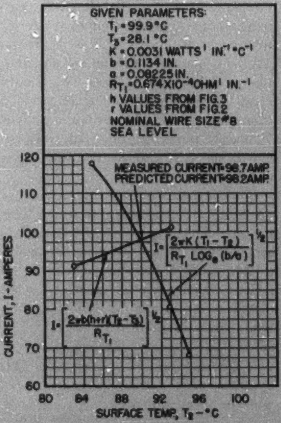


Figure 5. Graphical solution for the current satisfying the given conditions of heat transfer. Nominal wire size: AN 8

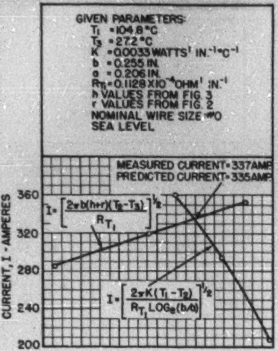


Figure 6. Graphical solution for the current satisfying the given conditions of heat transfer. Nominal wire size: AN 0

for the current as there are layers plus an additional equation derived from the relation for transfer of heat from the outer surface to the surrounding air and walls. These expressions can be solved for the current in a manner analagous to that described previously.

Alternative Method for Predicting the Current

A second procedure for determining the current necessary to produce a given conductor temperature now will be considered for the case in which the insulation consists of a single layer. The current required could be computed directly with equation 7 or 8 if T_2 were known. Thus, in equation 8 the unknown quantities are h, r, and T_2. When T_2 is known r can be obtained from Figure 2 and h from Figure 3. It will now be shown how T_2 may be predicted graphically with the aid of the heat transfer relations given earlier.

For the assumed conditions the rate of heat dissipation at a given atmospheric pressure is fixed when b, a, K, T_1, and T_3 are specified. The fixed heat dissipation establishes T_2, h, and r uniquely. The relations which must be satisfied by the eight quantities can be derived as follows. From equation 1

$$T_1-T_2=\frac{P\log_e(b/a)}{2\pi K} \tag{12}$$

while from equation 6

$$T_2-T_3=\frac{P}{2\pi b(h+r)} \tag{13}$$

Dividing equation 12 by 13

$$\frac{T_1-T_2}{T_2-T_3}=\frac{(h+r)b\log_e(b/a)}{K} \tag{14}$$

From equation 14 and the fact that

$$T_1-T_3=T_1-T_2+T_2-T_3$$

it follows that

$$\frac{T_2-T_3}{T_1-T_3}=\frac{1}{\dfrac{(h+r)b\log_e(b/a)}{K}+1} \tag{15}$$

Letting $(T_2-T_3)/T_1-T_3)$ equal (Q) and solving equation 15 for a gives the relation

$$a=b\epsilon^{-\frac{K}{b(h+r)}\left[\frac{1}{Q}-1\right]} \tag{16}$$

Conductor radius a now may be plotted as a function of b with equation 16 for specified heat transfer conditions, that is, for any set of K, T_1, T_2, and T_3 values. The required r and h coefficients are obtained from Figures 2 and 3. In Figure 8, two families of curves are shown. T_1, T_3, and K are fixed for each family; each of the three member curves is plotted for a different T_2. To increase the usefulness of the curves each member curve has been designated by the temperature ratio Q rather than by T_2. The computation details for one curve of Figure 8 are given in Appendix III. Figure 8 may be used to find T_2 for a cable of any conductor size and a wide range of insulation thickness where the values of K, T_1, and T_3 are those indicated. For example, Q for a cable whose a, b, K, T_1, and T_3 values are 0.100 inch, 0.150 inch, 0.0040 watt inch^{-1}, degrees centigrade^{-1}, 75 degrees centigrade, and 25 degrees centigrade respectively is found in Figure 8 to be 0.84. T_2 is, therefore, 67 degrees centigrade. The current for this case may now be computed with equation 8. Figures 9 and 10 are plotted for four additional sets of K, T_1, and T_3 values which are relevant to aircraft cable.

Figure 9 was used to determine Q for the three cables described in Figures 4, 5, and 6 under the same conditions given there. The agreement between the measured values and those obtained from Figure 9 is shown in Table I. The curves of this figure were extrapolated slightly to obtain the value of Q for the number 0 cable.

Table I

Nominal Wire Size	Q = $(T_2-T_3)/(T_1-T_3)$ Given by Figure 9	 From Actual Measurements of T_1, T_2, and T_3
18 (Figure 4)	0.81	0.81
8 (Figure 5)	0.85	0.85
0 (Figure 6)	0.83	0.83

Table II

Nominal Wire Size	Calculated Current Amperes	Measured Current Amperes
18	25	25.5
8	98	99
0	335	337

The current capacity of the three cables then was calculated with the aid of equation 8 and the surface temperature based on the predicted temperature ratios given in Table I. The calculated currents and those obtained by actual measurement are listed in Table II.

A complete discussion of the accuracy which can be attained by the two methods of predicting the current would require an analysis of the effect of each parameter appearing in equations 7, 8, and 16 on the estimated current. It may be pointed out that the effect of errors in each of the parameters is reduced by the square root which occurs in equations 7 and 8. An estimate of errors which are likely to be made in b, a, h, r, and K makes it reasonable to expect errors in current prediction up to 5 per cent.

Effect of Altitude on Current-Temperature Rise Relations

The measurement of the convection coefficients at pressures corresponding to 20,000 and 60,000 feet as well as at sea level makes it possible to determine the current needed to produce a given conductor temperature for altitude conditions. The two methods described previously are applicable. The convection coefficients for 20,000 and 60,000 feet for T_2-T_3 values of 30, 50, and 70 degrees centigrade are plotted as a function of over-all cable diameter in Figure 3.

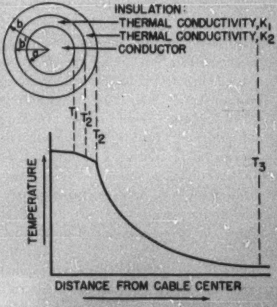

INSULATION:
THERMAL CONDUCTIVITY, K_1
THERMAL CONDUCTIVITY, K_2
CONDUCTOR
b
b'
a
T_1
T_2'
T_2
T_3
TEMPERATURE
DISTANCE FROM CABLE CENTER

Figure 7. Approximate temperature distribution for cable insulated with two layers

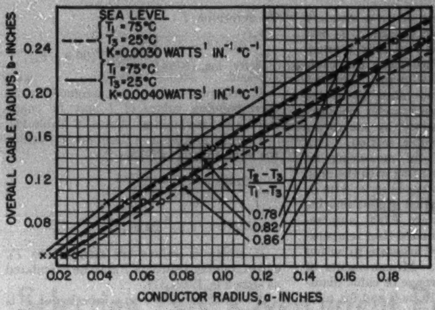


Figure 8. Conductor radius versus cable radius for specified heat transfer conditions

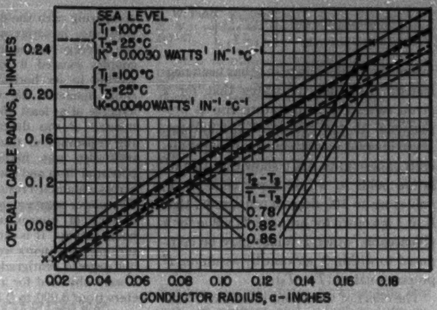


Figure 9. Conductor radius versus cable radius for specified heat transfer conditions

Curves similar to those in Figures 8, 9, and 10 may be drawn for both simulated altitudes.

While the current capacity of a given cable at a particular altitude is a function of the six parameters, T_1, T_3, b, a, R_{T_1} and K, the results of measurements on typical aircraft cables provide a good estimate of what may be expected generally. For eight cables, which included sizes from AN 20 to AN 0, the relative current values necessary to produce the same rise in conductor temperature are as shown:

Altitude in feet	0	20,000	60,000
Relative Current	100	93±2	82±2

Effect of Changing Cable Parameters

The effect of changing one or more of the cable parameters on the current capacity of a given cable is not self-evident because nine interrelated factors, a, b, K, T_1, T_2, T_3, h, r, and R_{T_1} are involved. A quantitative estimate can be made with the help of Figures 8, 9, and 10 and equation 8. The method by which the effect may be calculated will be made clear in two examples.

Let us suppose we wish to estimate the variation in current capacity when the thermal conductivity is changed. The effect of a change in K on Q can be observed by studying Figure 8. The two sets of curves plotted there were based on K equal to 0.0030 and 0.0040 watt inch^{-1} degrees centigrade^{-1}; T_1 and T_3 are the same for both families. Consider a cable whose conductor and insulation radii are 0.100 and 0.144 inch. When the thermal conductivity of the cable is 0.0030 watt inch^{-1} degrees centigrade^{-1}, Q equals 0.82. If the thermal conductivity is increased to 0.0040, Q becomes 0.86 as may be seen in the figure. The corresponding values of T are 64 and 66 degrees centigrade. The current capacities of the cable for both cases may be calculated by equation 8. The ratio of the two currents for these conditions, and, in fact, for any two conditions is given by

$$\frac{I'}{I}=\left[\frac{(R_{T_1})b'(h'+r')Q'(T_{1'}-T_{3'})}{(R_{T_1})'b(h+r)Q(T_1-T_3)}\right]^{1/2} \qquad (17)$$

where the primed and unprimed symbols each refer to one set of conditions. In the particular case being considered, equation 17 reduces to

$$\frac{I'}{I}=\left[\frac{(h'+r')Q'}{(h+r)Q}\right]^{1/2} \qquad (18)$$

since R_{T_1}, b, T_1, and T_3 equal $(R_{T_1})'$, b', T_1', and T_3' respectively. For the T_1 and T_3 values given in Figure 8, computation gives

$$\frac{I_{(K=0.0040)}}{I_{(K=0.0030)}}=1.03$$

An increase in K of 33 per cent has thus resulted in a 3-per-cent increase in current capacity. It can be verified by checking other points in Figure 8 that the effect of the change in thermal conductivity on the current capacity just described holds generally to a very good approximation.

The effect of changing T_1 can be seen by comparing points with identical coordinates in Figures 8 and 9. These curves show that when T_1 is increased from 75 to 100 degrees centigrade, Q decreases by approximately 1 per cent. The ratio of the current capacities for the two values of T_1 is found by substituting the required numerical values in equation 17.

Theoretically, the current ratio, I'/I, which results when T_1 alone is changed, would appear to depend on cable size. This dependence on size is due to the fact that the term $(h'+r')/(h+r)$ corresponding to a given pair of T_1 values will be different for different values of b. However, a survey of Figures 2 and 3 shows that the quantity $(h'+r')/(h+r)$, corresponding to the T_1 values, 100 and 75 degrees centigrade, is nearly constant for the range of b values in Figure 3. Therefore, the current ratio I'/I may be considered independent of cable size.

If we now observe that the change in R_{T_1} compensates for the change in $h+r$ when T_1 is increased, we find that equation 17 reduces to

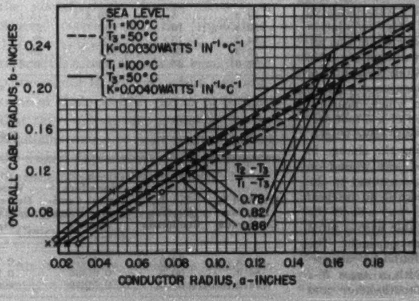


Figure 10. Conductor radius versus cable radius for specified heat transfer conditions

$$\frac{I'_{(T_1=100\text{ degrees centigrade})}}{I_{(T_1=75\text{ degrees centigrade})}}=\left[\frac{T_1'-T_3}{T_1-T_3}\right]^{1/2}$$

where the small change in Q has been neglected. Equation 19 states that for a given cable the temperature rise is proportional to the square of the current. Experimental measurements on eight aircraft cables up to AN 0 show that this relation holds to within 2 per cent for T_1 values up to 100 degrees centigrade, that is, if I is known for any one T_1 value for a given cable, other I values may be computed to within 2 per cent by assuming I^2 proportional to (T_1-T_3).

The effect of variations in ambient temperature T_3 and in the cable dimensions a and b can be studied in a similar manner.

The methods for determining current capacity and the effects of varying the cable parameters which have been described may be extended to any range of T_1 or T_3 values provided h is known as a function of b and T_2-T_3. The other parameters which are required, namely, R_{T_1}, a, b, r, and K, are either (1) independent of temperature or (2) may be calculated easily for the new range of temperature values. In addition, since the current rating can be computed for any set of R_{T_1}, b, a, and K values, modifications in cable design such as (1) use of conductor materials other than copper, (2) changed conductor and insulation dimensions, and (3) improved insulation add no difficulties to the cable rating problem.

Appendix I. Measurement of *h* and *K*

From equation 6 the convection coefficient can be written as

$$h=\frac{P}{2\pi b(T_2-T_3)}-r$$

To determine h experimentally, T_2, T_3, b, and P were measured while r was calculated with equation 4. In the latter calculation the surface emissivity of the cable was assumed to be 0.9. The heat-dissipating cable was suspended horizontally in the center of a closed chamber which measured 9 by 12 by 8 feet. The cable on which measurements were made consisted of a 5-foot section at the center of a 16-foot cable length. T_2 was found by measuring the change in resistance of a number 44 wire wound on the cable. The fine wire, cemented to the outer cable surface to make it an integral part of the cable surface, was wound the entire length of the test cable with a pitch of approximately three turns per inch. The resistance of the coil was found by measuring its potential drop with a type *K-2* Leeds and Northrup potentiometer and comparing with the drop across a standard resistance in series with the coil. T_3 was the average temperature of four thermocouples located 9 inches from the cable section in its horizontal plane. Over-all cable diameter, $2b$, was taken to be the average of at least 15 micrometer measurements. P was determined by measuring the resistance of the cable section. This resistance was found by measuring its potential drop and comparing with the drop across a standard resistance in series with it. The power source consisted of a submarine battery of approximately 6,000 ampere-hours rating. It is estimated that temperature measurements were accurate to within 0.5 degrees centigrade. Measurements of h were obtained for a range of cable diameters from 0.090 to 0.510 inch and for temperature differences, (T_2-T_3), extending from 10 to 80 degrees centigrade. Measurements were obtained for two ambient temperatures, 25 and 50 degrees centigrade for the sea level data. All altitude measurements were made in an ambient of 25 degrees centigrade.

The thermal conductivity of the cable insulation was obtained by measuring the quantities required by the relation

$$K=\frac{P\log_e(b/a)}{2\pi(T_1-T_2)}$$

derived from equation 1. P, b, and T_2 are measured as before, and a is an average of at least 15 micrometer readings. T_1 is determined from the measurement of the resistance of the cable section.

Appendix II. Computation Details for Figure 4

Given

$T_1=99.3$ degrees centigrade
$T_3=24.9$ degrees centigrade
$K=0.0035$ watt inch^{-1} degrees centigrade^{-1}
$b=0.0557$ inch
$a=0.0249$ inch
$R_{T_1}=0.000598$ ohm inch^{-1}

Table III

Assumed T_2, Degrees Centigrade	T_1-T_2, Degrees Centigrade	I^2, Amperes^2	I, Amperes
94.3	5	228.5	15.1
89.3	10	257	21.4
84.3	15	686	26.2
79.3	20	914	30.2
74.3	25	1142	33.8

To determine I

$$I=\left[\frac{2\pi K(T_1-T_2)}{R_{T_1}\log_e(b/a)}\right]^{1/2}=[45.7(99.3-T_2)]^{1/2}$$

If arbitrary values are assigned to T_2, corresponding values of I may be calculated and tabulated as shown in Table III.

I can be plotted as a function of T_2 using the values given in Table III.

$$I=\left[\frac{2\pi b(h+r)(T_2-T_3)}{R_{T_1}}\right]^{1/2}$$

$$I=[585(h+r)(T_2-24.9)]^{1/2}$$

If arbitrary values are assigned to T_2, corresponding values of I may be calculated as shown in Table III.

A second curve of I as a function of T_2 is obtained from the values shown in Table IV. The intersection of the two curves gives the desired current.

Table IV

Assumed T_2, Degrees Centigrade	T_2-T_3, Degrees Centigrade	h, Watts Inch^{-2} Degrees Centigrade^{-1} from Figure 3	r, Watts Inch^{-2} Degrees Centigrade^{-1} from Figure 2	h+r, Watts Inch^{-2} Degrees Centigrade^{-1}	I^2, Amperes^2	I, Amperes
94.9	70.0	0.01375	0.00500	0.01875	768	27.7
84.9	60.0	0.01330	0.00477	0.01805	634	25.2
74.9	50.0	0.01290	0.00455	0.01745	510	22.6

Appendix III. Computation Details for One Curve Plotted in Figure 8

Given

$T_1=75$ degrees centigrade
$T_3=25$ degrees centigrade
$K=0.0040$ watt inch^{-1} degrees centigrade^{-1}
$Q=(T_2-T_3)/(T_1-T_3)=0.78$

To calculate a as a function of b Substituting K and Q in

$$a=b\epsilon^{-\frac{K}{b(h+r)}\left[\frac{1}{Q}-1\right]}$$

we obtain

$$a=b\epsilon^{-\frac{1.128\times10^{-4}}{b(h+r)}}$$

which will now be evaluated for the given conditions over the range of b values needed for aircraft cable. In the present case

$T_2-T_3=0.78(T_1-T_3)=39$ degrees centigrade

r and h values will be selected for $(T_2-T_3)=39$ degrees centigrade from Figures 2 and 3 respectively. In addition, h will be chosen corresponding to the particular value of b indicated in Table V.

Conductor radius a now may be plotted as a function of b.

References

1. Heating and Current Capacity of Bare Conductors for Outdoor Service, O. R. Schurig,

Table V

b, Inches	h, Watts Inch^{-2} Degrees Centigrade^{-1}	r, Watts Inch^{-2} Degrees Centigrade^{-1}	$\frac{1.128\times10^{-3}}{b(h+r)}$	a, Inches
0.050	0.01295	0.0043	1.308	0.0135
0.100	0.00960	0.0043	0.812	0.0444
0.150	0.00815	0.0043	0.604	0.0820
0.250	0.0066	0.0043	0.414	0.1652

C. W. Frick. *General Electric Review* (Schenectady, N. Y.), volume 33, number 3, March 1930, pages 141–57.

1. How to Determine Current Capacity for Cables and Busses in Still Air, J. P. C. McMath. *Electrical World* (New York, N. Y.), volume 126, number 13, September 28, 1946, page 140.

2. The Calculation of Heat Transmission (book), Margaret Fishenden, Owen A. Saunders. His Majesty's Stationery Office, London, England, 1932.

No Discussion

Selecting A-C Overcurrent Protective Device Settings for Industrial Plants

F. P. BRIGHTMAN
ASSOCIATE MEMBER AIEE

FAULT current (also designated overcurrent or short-circuit current) protective devices, fuses, circuit breakers, and relays are the watchmen of a power system, whose job it is to detect trouble and get rid of it as expeditiously as possible. The design of such a protective system involves two separate although interrelated steps.

1. Selection of the correct devices to do the job.
2. Choice of correct and time settings for the adjustable devices that will enable them to function selectively to clear that portion of the system in trouble with as little disturbance to the rest of it as possible. The two steps are interrelated in that the devices selected for a given system must be capable of the required range of current and time settings needed or inherently have the desired characteristics incorporated in their design.

On all but the simplest systems there will be two or more circuit breakers, or other circuit interrupting devices, between a fault and the source of power. In order to localize the disturbance as much as possible, these devices should be selective in operation, so that the one nearest the fault on its power-source side will have the first chance to operate. If for any reason this protective device fails to function on schedule, the next device in the chain must be ready to take over the job of opening the circuit, and so, in successive steps to the power-source circuit breaker itself, if necessary, to clear the fault. To accomplish this objective, the fault current protective devices must have been selected, as in the case of fuses and instantaneous trips, and so forth, or be capable of adjustment to operate on the minimum current that will permit them to distinguish between fault current and permissible load-current peaks, and to function in the minimum possible time and still be selective with others in series with them. When these two requirements are met, the damage to equipment, or the interference with production due to loss of power during a short circuit, or both, will also be at a minimum.

On the very simple systems with but one fault current protective device between the power source and the load there is no co-ordination problem, but it is still necessary to choose a current setting compatible with the load characteristics.

All of the adjustable devices must be set in the field to achieve the desired co-ordination. The purpose of this discussion is to describe in detail some of the basic fundamental procedure involved in making the necessary study of an a-c power system to determine what the current and time settings of the adjustable devices should be, assuming that they and the nonadjustable devices have been selected correctly for the application. There is no mystery involved in doing so, rather it is a case of patient, even dogged perseverance in trying various combinations of characteristic curves to be sure that they will operate positively on the minimum values of fault current and in the desired sequence on all currents up to the maximum currents expected.

Individual Protective Device Characteristics

An overcurrent protective system is simply a multiplicity of co-ordinated individual devices—fuses, direct-acting trip coils, or plunger and induction-type relays. Therefore, an understanding of their individual behavior, in so far as it affects their co-ordination with other devices, is a logical starting point for a general discussion of protective device settings.

FUSES

Since the time-current characteristic of fuses is nonadjustable, it is necessary to take cognizance of their time-current characteristics when they are initially selected for the installation. When doing so, it should be noted that the manufacturer's time-current curves may be plotted in terms of melting time, minimum melting time, or total clearing time. Melting time is an average value of melting time from tests, whereas minimum melting time is the shortest time at which the particular type and rating of fuse will melt for a given value of current through it. Total clearing time is the maximum expected time required for the fuse to melt and the arc to clear, see Figure 1.

Unfortunately, there is no widely accepted industry-standard method of plotting fuse characteristics, and arc clearing time varies with the type and rating of fuse. Consequently, it is necessary to obtain specific data from the fuse manufacturer if accurate data is required.

Generally speaking, however, the tolerance band for manufacturing variations, and so forth, will be ±10 per cent in current for melting-time (average) fuse curves and +20, −0 per cent in current for minimum time curves. It is also considered good practice when co-ordinating fuses to allow an additional tolerance of −25 per cent in time rather than current.

Paper 52-184, recommended by the AIEE Industrial Power Systems and Relays Committees and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted January 30, 1952; made available for printing May 6, 1952.

F. P. Brightman is with the General Electric Company, Schenectady, N. Y.

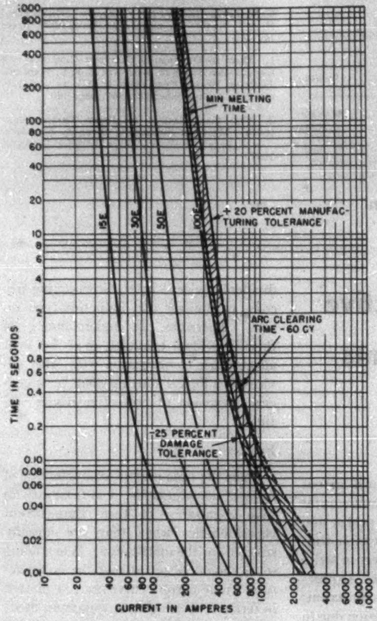


Figure 1 (left). Melting time curves of current limiting fuses

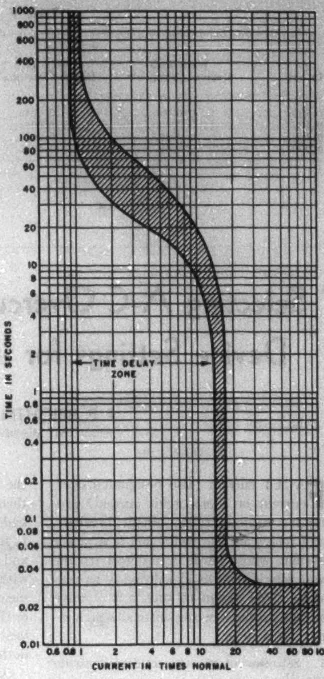


Figure 2 (right). Typical time-current characteristic of 600-volt air circuit breaker with long-time and instantaneous trip mechanism

The latter provides a time cushion to insure against possible damage to the fuse while waiting for some other protective device in series with it on its load side to clear the fault. The total operating time band is simply the minimum melting time, plus the −25 per cent damage tolerance, the +20, −0 per cent manufacturing variations tolerance, and the arc clearing time. Since the latter is of very short duration, it can be neglected safely when co-ordinating fuses with relatively slow acting devices, such as inverse time induction relays and time delay trip coils.

The minimum melting-time curve plus the damage-tolerance factor should be used when the fuses are ahead (nearer the source) than the other device with which they are being co-ordinated. Conversely, total clearing time must be considered when the fuse is on the load side of the other device.

Instantaneous Direct-Acting Trips or Relays

Strictly speaking, such devices are not quite instantaneous in operation on current magnitudes near their pickup setting, but for all practical purposes they can be considered as such. Possible exceptions would be encountered in the case of very close co-ordination of such devices with fast blowing fuses. Some of these instantaneous devices are adjustable as to current pickup, but whether adjustable or nonadjustable they will operate in essentially the same time regardless of current magnitude, as long as it equals or exceeds their pickup value. Current adjustment, when possible, is accomplished by raising or lowering the armature of direct-acting trips or plunger-type relays.

Time-Delay Direct-Acting Trips

Time-delay direct-acting trips will be encountered occasionally on the operating mechanism of power circuit breakers (over 600 volts) and almost always on low voltage (600 volts and below) circuit breakers. They can be adjusted for current pickup in the same manner as the instantaneous trips. The time-delay characteristic is usually obtained by means of a liquid-filled dashpot, an escapement mechanism timer, an oil-film timer, or a thermal trip. The time delay of some low-voltage circuit breakers can be adjusted in the field, but many of the modern designs are preset at the factory in accordance with the time-current characteristics presumed to be best suited for the application. Either type has a considerable tolerance band to allow for manufacturing variations, liquid viscosity variations due to temperature, and arc clearing time. A typical time-current curve of a 15,000-ampere interrupting capacity, 600-volt circuit breaker with an instantaneous trip is shown in Figure 2. The current is plotted in terms of multiples of current pickup. In this case the coil is set to pick up at 100-per-cent rating, but it can be adjusted over a range of 80 to 160 per cent of the trip coil rating. It will be noted that both the time delay and instantaneous device pickups have a tolerance of ±10 per cent. The time-delay portion of these devices will continue to move (overtravel) after the coil has been de-energized, as many other protective devices do, and allowance for this tendency is included in the tolerance band of the time-delay portion of the curve.

Since the tolerance band incorporates all the necessary allowances including arc clearing time, these devices will be selective as long as their plotted tolerance bands are entirely separate from each other.

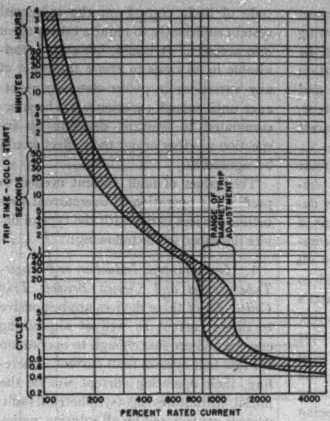


Figure 3 (left). Typical time-current curve of 100-ampere frame size (70–100-ampere range) molded case circuit breaker with thermal magnetic trip

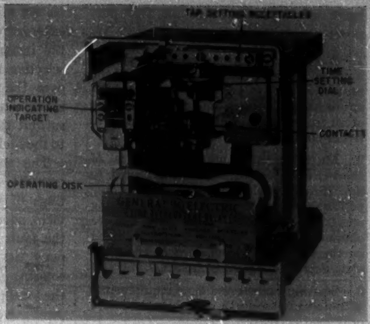


Figure 4 (right). Standard inverse induction relay

THERMAL MAGNETIC TRIPS

This type of trip, which depends on a thermal device for its time-delay operation, is used on some conventional low-voltage air circuit breakers and on the so-called molded-case type low-voltage circuit breakers. A typical time-current curve including necessary tolerances and arc clearing time for such a device with an instantaneous magnetic trip is shown in Figure 3. These devices will also coordinate as long as their plotted tolerance bands do not touch each other, same as discussed in the preceding section.

INDUCTION RELAYS

Induction-type relays are generally used on modern switchgear over 600 volts and to some extent on low voltage (600 volts and below) where greater accuracy is required than can be provided by the direct-acting time-delay trip coils on the circuit breakers. There are many variations of induction-type relay, but all of them stem from the standard induction-type, inverse-time overcurrent relay pictured in Figure 4. The operating coil has several current-range taps which are connected to a tap plate as shown. The desired pickup current is obtained by inserting the tap plug in the receptacle to which that tap is connected. The time setting is adjusted by means of a dial or lever (mostly on older models). Moving the dial or lever varies the distance which the disk has to travel to close the relay contacts. The time dial or lever can be set at any point on the scale, which is arbitrarily calibrated uniformly from 0 to 10. A family of operating time-current curves for a typical inverse-time induction relay is shown in Figure 5. The curves are plotted in terms of time, and multiples of minimum closing or pickup current. Pickup is that current at which the relay will just start to move, and multiples of it equal current-transformer secondary amperes divided by relay-tap amperes. These relays have several taps, for example, 4, 5, 6, 8, 10, 12, 16 for the standard 4–16-ampere relay.

To illustrate the use of these curves in determining relay performance, let us assume: 1. a 400/5-ampere current-transformer (CT) ratio, 6,000-ampere fault current, relays set on the 10-ampere tap and 1.5 time dial required to find the operating time; 2. a CT ratio of 500/5 amperes, maximum load current 450 amperes, short-circuit current of 8,000 amperes, desired relay operating time 0.5 second, what time-dial setting is required.

In case 1 there will be 75 amperes in the CT secondary with 6,000-ampere fault current in the primary. This is 7.5 times the specified 10-ampere relay tap or pickup setting. Now entering the Figure 5 curves at 7.5 multiples of minimum closing current and proceeding vertically to an interpolated 1.5 time-dial curve, we find that the relay operating time would be 0.55 second.

In case 2 the current in the CT secondary corresponding to the load current of 450 amperes will be 4.5 amperes. Since this is to be the maximum normal current in the circuit, the relay can be set on the 5-ampere tap. The primary fault current of 8,000 amperes produces 80 amperes in the CT secondary, which is 16 multiples of the selected 5-ampere tap setting. Now if we enter the relay time-current curve, see Figure 5, at 16 multiples of minimum closing current (16 times pickup) and proceed vertically to a point representing 0.5 second, we find that the relay would have to be set at 2.1 time dial.

TIME-DELAY PLUNGER-TYPE RELAY

Relays of this type may also be encountered in making relay studies. Their current pickup setting is adjusted by raising or lowering the magnetic plunger in the calibrated plunger-tube. Timing depends on the rate at which air escapes from a bellows at the top of the relay and is varied by turning the air valve, which is also calibrated from 0 to 10 in some modern designs.

RELAY SETTING—MISCELLANEOUS GENERAL FACTORS

The relays and other devices previously described are those that will be encountered most often on industrial power systems. In addition, numerous other modern overcurrent plus differential and other type protective devices, as well as many obsolete models, will be found. Basically, different manufacturer's versions of a given type of relay or tripping device will conform to the general patterns discussed. Nevertheless, there are sufficient variations in time-current curves to necessitate procurement of the manufacturer's data for the particular device involved, if reasonably accurate settings and performance are to be obtained.

Data Required for a Relay-Setting Study

Now that we know in general what to expect in the way of operating time performance from the various overcurrent protective devices on a system, the next problem is to secure the necessary data from which specific relay or other overcurrent protective device settings may be determined. Following is a list of the basic information needed for this purpose:

1. A 1-line diagram of the power system involved, showing the type and rating of the protective devices and their associated instrument transformers and the impedances of all transformers, rotating machines, and feeder circuits.

2. Maximum and minimum values of short-circuit current that are expected to flow through each protective device whose performance is to be studied under varying operating conditions. This data can be obtained from a short-circuit study based on the information contained in the complete detailed diagram of item 1.

3. Starting current and time requirements of motors, and so forth, and their maximum peak-load current.

4. Manufacturers' characteristic performance curves of the relays, trip coils, and fuses to be co-ordinated.

5. Manufacturers' performance curves of current transformers, especially bushing types.

6. Any special overcurrent protective requirements such as those stipulated by the National Electrical Code or dictated by the load characteristics.

7. Any special overcurrent protective device setting requirements stipulated by the public utility company with which the industrial plant may be interconnected.

8. Decrement curves showing the rate of decay of the fault current supplied by generators.

Short-Circuit Current Calculations for a Relay Study

The basic data and procedure required for making short-circuit calculations as a preliminary to a relay study is the same as for a circuit-breaker duty study, but the combinations studied will be somewhat different. In a relay study the problem is to determine the characteristics of devices that: 1. will be sure to operate on the minimum values of fault current expected at certain fixed times following the instant of short circuit, and 2. be selective in their sequence of operation over the range between minimum and maximum values of short-circuit current, so that the relay or other device nearest to the fault will be the first to operate.

The magnitude of the short-circuit currents, which will determine the settings of the overcurrent protective devices, should be calculated on the basis of the fault current from any power company ties, plus that contributed by all rotating machines directly connected to the local power system, that is, the user's own power generation and distribution system. It is, of course, unlikely that every one of these machines will be in operation simultaneously under normal conditions, but it could be the case during transfer periods when load is being switched from one machine to another, hence the relays, and so forth, must be capable of giving the desired sequential operation on that magnitude of current.

Usually the fault current contribution from the power company can be calculated on the basis of a single reactance value because the relatively high reactance of the tie circuit supplying the industrial plant is alone sufficient to limit the fault current so much that the effect of differences between maximum and minimum generating capacity and subtransient ($X_d''d$) and transient (X_d') reactances are unimportant. Also the decrement effect will be eliminated for all practical purposes, see curve *4*, Figure 7. If, however, the tie circuit between power company and industrial is of large kilovolt-ampere capacity and relatively low reactance, variations in power company system reactance will affect the magnitude of fault current on the industrial system sufficiently to require taking them into account. It would, however, be advisable in all cases to check the effect of any variations in power company reactance until enough experience has been accumulated to warrant deciding by inspection whether or not their effect would be significant.

The values of fault current needed to predict the operation of overcurrent protective devices varies with the type of device described in the following:

1. *Instantaneous Induction or Plunger-Type Relays and Circuit Breaker Trip Coils.* All of these devices are responsive to direct current as well as alternating current and are fast enough to operate on the first half-cycle of fault current. Therefore, their operating current will be the initial (instantaneous) asymmetrical fault current contributed by all rotating equipment (including induction motors and generators) calculated on the basis of its subtransient reactance (X_d''). Use a 1.6 multiplier to obtain the magnitude of the asymmetrical or offset current, except on systems of 5 kv and below when there are transformers or appreciable line resistance between the point of fault and the source, in which case use a 1.5 multiplier.

2. *High-Speed Induction-Type Relays.* These relays operate in 3 cycles or less and are also responsive to d-c components of current, consequently the initial X_d'' asymmetrical fault current from item 1 may be used as their operating current also. Actually, the current magnitude will drop during the 3 cycles due to the decrement effect, but the influence of this reduction can generally be neglected in industrial system relay studies.

3. *Time-Delay Relays and Trip Coils.* Generally speaking the devices of this class are too slow to be appreciably affected by the subtransient (X_d'') reactance values of fault current. Therefore, their maximum operating current will be the initial value of symmetrical transient (X'_d) reactance current supplied by the power company and the local plant's synchronous equipment (generators, condensers, and motors). Current-time settings of such devices that permit operation in approximately 0.1 second (6 cycles) on high-current magnitudes are theoretically an exception to this rule, but practically, such relays usually will be the first in a series of relays (farthest from the power source), so that any reduction in their operating time simply increases the time margin between that relay and the one ahead of it. When the operating time of

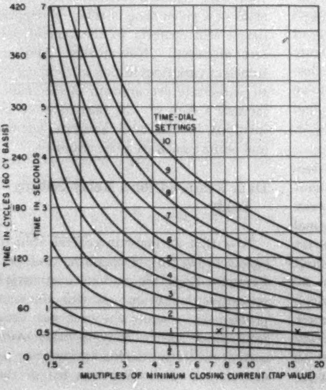


Figure 5. Time-current curves of standard inverse induction relay

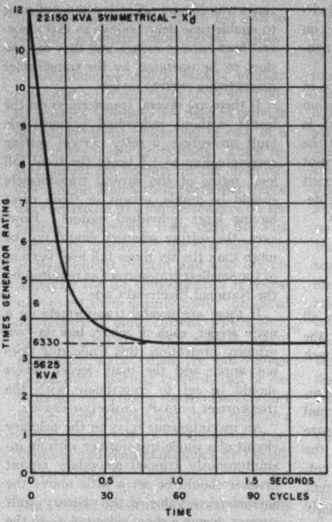


Figure 6 (above). Short-circuit decrement curve of 1,875-kva 3,600-rpm 2,300-volt generator with regulator in use and exciter field rheostat set to limit exciter ceiling voltage to 120-per-cent rated volts

relays is desired with greater accuracy, such as for co-ordinating with generator overcurrent relays with voltage restraint, it is necessary to construct a decrement curve of the generators supplying power during that particular operating condition, and then determine how the relay will perform on the decaying fault current. The need for decrement curves in determining fault currents will be discussed in greater detail later.

4. *Fuses.* As stated previously, the performance of fuses is an inherent characteristic of the particular unit involved and must be taken into account when they are selected for the job. Nevertheless, the following discussion of expected fuse performance seems desirable at this point to afford a better understanding of the problem of co-ordinating them with other devices.

Fuses of all types depend on the heat created by the passage of the fault current through them to melt their fusible elements. Since the heat is due solely to the I^2R loss, d-c as well as a-c components of current will affect them, so the initial asymmetrical X_d'' reactance current will affect fuse operation. However, the effect of the d-c offset current decreases in importance very rapidly as the expected fuse melting time increases.

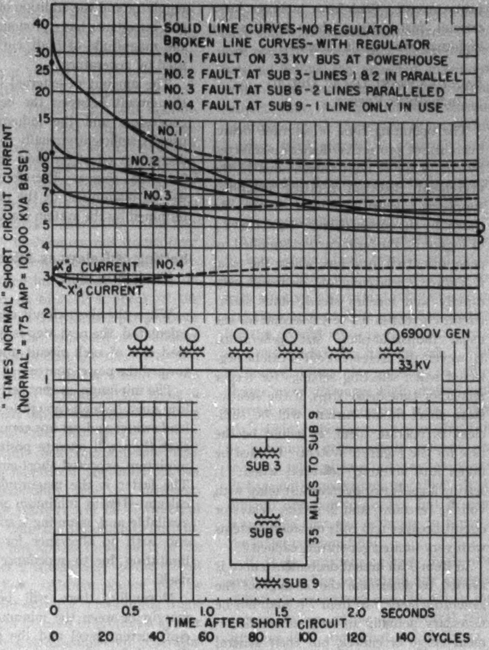


Figure 7. Family of short-circuit decrement curves on 33-kv system supplied by 2 10,000- and 4 5,000-kw turbine-generator transformer units

The melting time of fuses varies from less than 0.01 second to several hundred seconds, depending on the fault-current magnitude. For example, the *100E* current limiting fuse in Figure 5 has a melting time of 0.01 second (0.6 cycle) at 2,700 amperes, 0.1 second at 780 amperes, 10 seconds at 320 amperes, and 300 seconds at 200 amperes. For good performance on a system which has most of its power locally generated, the fuse minimum melting-time curves should show that the initial magnitude of unrestricted (no intentional limitation) fault current under minimum power-supply conditions is sufficient to blow the particular fuse within 1 or 2 cycles (0.033 second). If they cannot do that, the use of fuses should be carefully scrutinized because the rapid generator decrement effect may reduce the fault current so much that the fuses either might not blow at all or else might be objectionably slow. Also there would be no margin to allow for any current limitation due to fault impedance. An exception to the latter statement would be the case of a fuse at a location where the fault current was reduced substantially by high circuit impedance as shown in curve *4* of Figure 7. Under such a condition the difference between the initial symmetrical X_d'' reactance current as represented by the curve and the initial X_d' value of current as represented by the dot is negligible, and so is the decrement effect. Therefore, if the X_d' current is sufficient to blow the fuse in a few cycles, the application is satisfactory.

Modern power systems are often designed to have quite low values of line-to-ground fault current. In such cases, it will be found that the available ground fault current, particularly under minimum generating-capacity conditions, will be so low as to make large fuses either extremely slow or inoperative. When such a condition is encountered, additional means such as relays and circuit breakers must be provided for opening the circuit under line-to-ground fault conditions.

There are certain types of low-voltage fuses that are purposely designed to be slow on low magnitudes of fault current in order to ride through high normal load peak currents. Such fuses, however, have normal fast-melting characteristics on high-magnitude fault currents.

FAULT CURRENT DECAY DUE TO DECREMENT EFFECT

The operating time of all time delay, overcurrent protective devices is affected by the decay of fault current due to the decrement effect of rotating machines supplying short-circuit current. As shown by the curves (Figure 6 and curve *1* of Figure 7), the fault current supplied by generators to a local bus, where the total reactance is relatively low, decreases very rapidly and reaches steady-state short-circuit current in 0.5 to 1.0 second for the usual industrial-plant size generators. However, the decrement effect can be neglected when selecting settings for feeder relays, or time-delay trips, if the steady-state short-circuit current will be sufficient to operate them, as would be the case for the Figure 7 system. It must be taken into account, however, when: 1. setting a generator overcurrent relay with voltage restraint, and 2. setting relays or circuit breaker trip coils on small systems with very limited generating capacity.

In item 1 an actual decrement curve is needed to determine the settings of the generator relay. In item 2 it may not be necessary actually to construct a decrement curve or curves, but their general behavior and effect on the magnitude of fault current must be visualized mentally and taken cognizance of to be sure that the protective devices will operate on the current available at the time they are supposed to function.

For example, the decrement factor must be given consideration in those cases where a relatively small generator unit supplies power to a large system over weekends. On such a system it is quite possible that the available fault current will be insufficient to operate the relays on large feeders, in which case a practical solution would be to let the generator relay trip the generator circuit breaker and shut down the whole system in case of trouble on such feeders. Such a general shutdown can usually be tolerated during weekends when no essential load is in operation.

It is particularly important to keep the decrement effect in mind when selecting the devices, as well as their settings, for a system supplied from one or two engine-driven or geared-turbine-driven, salient-pole generators, because even their initial fault current is quite low due to their high inherent reactance.

It will be noted from the curves in Figure 7 that the presence of a voltage regulator and the addition of external circuit impedance has a pronounced effect on the magnitude of fault current contributed by generators.

The amount of external impedance in the circuit between the central station generators and the industrial plant is usually sufficient virtually to wipe out the decrement effect on short-circuit current from that source. Therefore, the current will be maintained at the magnitude determined by the system reactance.

Setting of Relays as Affected by Their Function

When the short-circuit currents are all calculated, the next step is to start at the load end of each circuit and work back toward the power source.

The minimum current setting of the relays must be high enough to carry normal load swings and yet low enough to be sure that they will operate positively on the minimum expected short-circuit current. The latter is the magnitude of current calculated with minimum power supply available and assuming a solid short-circuit with no allowance for any current limitation due to impedance in the fault itself.

Normally, there will be plenty of margin between the minimum short-circuit current level and the relay setting dictated by the maximum permissible load. Occasionally, however, that will not be true, such as in the case of high-reactance generators whose total output is transmitted over a single line. Such possibilities must be recognized and provided for initially when designing the protective system by specifying devices that do not depend on current magnitude alone for operation and selectivity.

The following general rules will serve as guides in determining the minimum relay current settings that will not trip on permissible load currents.

Relays on Transformer Feeders. The National Electrical Code states that time-delay overcurrent relays on a feeder for a single transformer should be set not more than 2.5 times the full load rating of the transformer for units without secondary circuit breakers, and up to six times transformer full load current if there is a secondary circuit breaker and the transformer reactance is around 0.06 per unit. The fact that the Code permits it, however, does not mean that the relays always should be set at these upper limits. If the load on the transformer is diversified with no relatively large motor units whose starting current might be the limiting item, a relay current setting of 1.5 times transformer rating is often sufficient to enable time-delay relays to carry normal load-current swings, and they are too slow to be operated by the transformer magnetizing current inrush.

If there are several transformers on the feeder without individual primary-side fault protection, a relay current setting corresponding to 1.5 times the total full load rating of the several transformers should be ample, except in those cases having large individual motors. However, the setting selected should not be more than the six times full load current of the smallest transformer as required by the National Electrical Code.

If there are several transformers on a main feeder, each of which has its own primary protection, this Code rule does not apply and the main feeder relays should be set to co-ordinate with the transformer primary protective device.

An instantaneous relay in the primary circuit of a single transformer with an instantaneously tripped secondary circuit breaker should be set a little above the asymmetrical value of the primary fault current during a short circuit close to the transformer secondary terminals. This setting is usually high enough to clear the magnetizing current inrush of the transformer, although the latter theoretically may be as much as 15 times normal for a few cycles. If the primary feeder circuit breaker trips when the transformer is energized, the relay current setting should be increased until the circuit breaker will stay closed.

The minimum setting of the instantaneous relay in a high-voltage feeder with several transformers on it will quite likely be dictated by the need for riding over their total magnetizing current inrush. As stated, this will be in the neighborhood of 10 to 15 times normal, which should be high enough to clear any probable condition such as a short circuit on the secondary of one or two units and normal load current on the others.

RELAYS ON SINGLE-MOTOR FEEDERS

The minimum setting of a time-delay relay intended to provide short-circuit protection for a motor feeder has to be above the maximum motor starting current so that the relay will not operate on the latter. This is necessary because any time-delay overcurrent relays that are expected to afford worth-while fault protection are much too fast to ride through motor starting current. The long time-delay induction relays that can be set lower than this in current and still be able to ride through the starting period afford a measure of normal overload (overheating) protection, but are much too slow to provide adequate short-circuit protection.

Instantaneous relays, usually plunger type, with a current range of 10–40 or 20–80 amperes, normally are used to provide fast short-circuit protection in case of a fault in the motor or its leads. These relays should not operate on the short-circuit current which their particular motor will contribute to a fault elsewhere on the system. Therefore, since they are fast enough to operate on the first half-cycle of short-circuit current and also are affected by the d-c component, it is necessary to set them slightly above the initial asymmetrical fault current which their motor can supply. This current, which is slightly higher than the full-voltage locked-rotor starting current, can be determined by dividing the full load current by per unit subtransient (X_d'') reactance of the motor (either synchronous or induction) and then multiplying by the offset factor of 1.6.

Relays on Incoming Lines and Feeders with Miscellaneous Load

The minimum setting of a time-delay overcurrent relay on an incoming line or feeder with miscellaneous load, which may include one or more motors, should be just above the expected peak load on the circuit. This usually will be the total of the starting current of the biggest motor plus full load on the other load. This will hold unless it is intended to start more than one motor at a time normally, or the motor controls are so designed that the motors would restart automatically when the circuit is re-energized following an outage.

Instantaneous relays may not be usable on such circuits because of inability to co-ordinate them with other instantaneous relays in the branch circuits as previously discussed. If they can be used, they should be set a little above the total of initial asymmetrical short-circuit current which all the motors could contribute to a short circuit elsewhere on the system, plus full load current on the nonmotor load. It is unlikely that any normal load surges would exceed this value.

Residually Connected Ground-Fault Relays

Residual ground-fault relays are connected in the Y of the secondaries of three current transformers where they will see only the unbalanced residual current flowing during line-ground faults. Under normal balanced 3-phase load conditions, no current will flow in the residual relay. Therefore, it can be set at low current levels to give sensitive operation on ground faults, which incidentally are the most common variety. Such relays are normally rated 0.5–2.0 amperes, or 1.5–6.0 amperes. These relays will operate satisfactorily over their entire tap range when connected to high-accuracy current transformers, but they may not on the 0.5-ampere tap when connected to the low turn ratio taps of some bushing current transformers. This is due to the fact that the bushing-current-transformer ratio breaks down due to the high volt-ampere burden imposed by the relay on the low-current tap. If a check of the bushing-current-transformer performance curves with the expected volt-ampere burden indicates that the relay may not operate, it will be possible usually to obtain satisfactory operation by using a higher current tap on the relay or bushing current transformer, or both.

Ground-fault relay current and time settings are handled in the same manner as for phase-fault relays. However, ground-fault current will be of different magnitude than the phase-fault current, and it may not appear in all parts of the system since a transformer in the circuit blocks its flow. Also its source may be different; for example, a grounding transformer might be used to ground a system supplied by a delta-connected generator.

Frequently ground-fault relays can be set to operate faster than the phase relays at the same location because the latter must co-ordinate with other phase overcurrent protective devices on the secondary of transformers for example, whereas a residual-ground relay on the primary will not see ground-fault current in the secondary system.

Residually connected, ground-fault relays are subject to possible false operation due to the CT error currents during a phase fault. This possibility is caused by the fact that duplicate current transformers are not exactly alike, even though made from the same batch of material. Although the variations are small, they are enough to cause appreciable amounts of unbalanced current to flow in the Y of residually connected CT's during severe phase-phase or 3-phase short circuits. Such unnecessary operation of the ground relay is of minor importance if the ground and phase relays involved control the same circuit breaker, but it might be quite troublesome if the ground relay was on a main-line circuit breaker whose tripping would shut down several branch feeders when only one of the latter was in trouble.

One solution to the error-current tripping problem is to design the system to have sufficient ground-fault current to insure positive tripping with the ground relays set above the expected maximum center-tap error current. Opinions on the subject vary, but ground-fault current approximately 5–15 per cent of maximum 3-phase fault current is generally considered sufficient. Conformation with this practice means ground-fault current of several thousand amperes and consequent equipment damage.

Another and perhaps preferable way is to keep ground-fault current and relay-current settings at a minimum consistent with positive low-current relay operation, and then prevent false operation by timing the ground relay to be selective with any phase relays on its load side. This, of course, raises the question of which causes more damage, high ground-fault current and less sensitive relay pickup with slightly shorter relay operating time or minimum ground-fault current and sensitive relay pickup but maybe one step (0.4 second) slower operating time. It is a difficult question to answer, but it seems reasonable that the odds are in favor of the latter combination.

If there are single-phase loads connected from line to neutral, any expected unbalance in current must be taken into account when setting the neutral relay, whether it is connected residually or in series as will be discussed here.

Ground-Fault Relays in Series with Generator or Transformer Neutral

It is fairly general practice to connect a final backup ground-fault relay to a CT in the generator or transformer neutral earth connection. A low-current relay and a CT whose current rating is 25 or 30 per cent of that of the neutral grounding resistor for that unit, is the combination generally used. This combination, which affords maximum sensitivity and protection, is permissible since current flows in the neutral of a balanced, 3-phase 3-wire circuit only during a line-ground fault.

The relays usually furnished for this purpose are capable of being set slow enough to permit the flow of ground-fault current up to the thermal limit of the neutral grounding resistor, if there is one, and many users set neutral relays to do that. Since it is fairly general present-day practice to use 10-second rated resistors, the relay would be set for 7–10 seconds in such cases. This slow setting is probably a carryover from former days when it was standard practice to use a 60-second or longer rated resistor. The accompanying relay system would first sound an alarm immediately in the hope that the operator might be able to clear the fault soon enough to avoid system shutdown (manually trip a feeder circuit breaker which had failed to open, for example), and later trip the main generator or transformer circuit breaker in time to avoid dangerous overheating of the resistor. In actual practice it is doubtful that the operator was able to do much in the 60 seconds allowed, and he could accomplish even less in the present-day 10-second time allowance.

Another method of setting the neutral relay is to time it to co-ordinate with the other system ground relays, that is, one time step slower. This seems preferable, since experience indicates that if a relay or circuit breaker trip coil is going to work at all it will do so in its normal time, which means that any further delay simply means more damage caused by the ground fault.

Co-ordination of Relays in Series

TIME-DELAY RELAYS

As mentioned previously, the time-current curves of direct-acting time-delay trips, fuses, and time-delay thermal devices include the necessary allowance for overtravel, and manufacturing tolerances, and so forth. The time-current characteristics of relays, on the other hand, are represented by families of single-line curves (Figure 4) to which tolerance bands must be added. Most relay time-current curves begin at 1.5 multiples of minimum closing current or pickup setting because their performance cannot be predicted too accurately below that value. However, curves showing expected time-current performance down to 1.1 times pickup usually can be obtained if required.

This time margin or tolerance band is based on the fact that the second relay in a chain continues to see fault current until the circuit breaker associated with the first relay has opened and the arc has been extinguished. This is nominally 8 cycles for the circuit breakers commonly used in industrial systems, although actually the opening time will be 4–5 cycles. Then after the first circuit breaker has opened the circuit and de-energized the second relay, the latter's contacts will continue to close for 0.1 second (standard inverse relay) due to the inertia in the induction disk of the relay to which the movable contact is attached.

A minimum total time margin of 0.4 second with maximum fault current flowing should be sufficient to afford satisfactory selectivity between standard inverse relays. This margin allows for the 0.13-second circuit breaker opening time (8 cycles), 0.1-second overtravel, and a safety factor of 0.17 second to cover manufacturing variations and inaccuracies in positioning of the time dial or lever when setting the relay.

The minimum time margin between very inverse and extremely inverse relays can be reduced to about 0.33 second because their over-travel is less and a shorter safety factor is needed since the difference in relay operating time caused by changing from one time-dial setting to another is also shorter on these higher speed relays, so that the positioning error is reduced.

If the total accumulation of even these short time intervals makes the operating time of the last relay (one nearest the source) undesirably long, the total time can be reduced a little by very careful checking of relay operating time. Generally speaking, however, the fraction of a second saving that might be made in the time margin between these values and the minimum possible settings that would not cause false operation, is not of sufficient importance on industrial power systems to warrant the extra trouble involved in making settings.

When selecting current-tap and time-dial settings for induction relays, it should be borne in mind that when two relays in series are set to co-ordinate properly under the maximum value of fault current expected, they always will be satisfactorily selective on lower values of current if they have the same shape time-current curves and the current setting of the slower relay is equal to, or higher than, that of the faster relay. If the current setting is lower, the operating-time curves of the two relays will cross each other at some low value of fault current, and the slow relay will beat the fast one for all currents below that value.

Another factor to be considered, when choosing between two combinations of current-tap and time-dial settings, either of which will give a desired operating time with maximum fault current flowing, is that the combination with the lower current and higher time-dial setting is usually preferable. The reason is that the relay with such a setting will be more sensitive and faster on low values of fault current. For example, suppose an operating time of 0.5 second is desired with a relay connected to 1,000/5-ampere CT's in a circuit with maximum symmetrical fault current of 20,000 amperes. Relays with 6-ampere tap and 2.1 time-dial or 10-ampere tap and 1.7 time-dial settings will both give the desired time. But in case of a fault involving only 3,000 amperes, the relay with the 6-ampere setting would operate in 1.25 seconds compared with 2 seconds for the 10-ampere combination. If the current is still further reduced to 2,000 amperes, the first relay will still operate in 2.1 seconds, but the second one will be very, very slow, since the current is only 1.0 times relay pickup at which point operation is uncertain.

INSTANTANEOUS RELAYS

When two instantaneous relays are used in series, their selectivity is dependent solely on their current settings. Therefore, they must be set so that the second relay will not pick up on the maximum asymmetrical current which the first relay (nearest the source side of the fault) can see. Such being the case, there must be sufficient impedance in the circuit between instantaneous relays in series to create the necessary current differential. If this objective cannot be attained, the second relay must be rendered inoperative by short-circuiting its coil.

Usually the impedance of a transformer is sufficient to permit co-ordinating an instantaneous relay on the high-voltage feeder panel with the instantaneous trip-coil of the low-voltage secondary circuit breaker. Also, the reactance of open transmission lines is frequently sufficient to provide the necessary differential in short-circuit current magnitude to permit the use of instantaneous relays at both ends.

Instantaneous relays cannot be co-ordinated on ordinary-length cable systems, because the circuit impedance is too low to cause the necessary current differential. In such cases, the relay at the bus end of the feeder should be utilized and the other one made inoperative to afford fast protection to the cable as well as the load.

Incidentally, the fact that a relay-setting study reveals that some of the instantaneous relays must be made inoperative should not be interpreted as a sign of a poorly designed protective system. This is so because it is quite common practice to include instantaneous attachments on all of the time-delay overcurrent relays on switchgear equipment so that they will be interchangeable.

Summary

The preceding discussion presents the basic data and factors involved in making a so-called relay study to determine the proper time-current characteristics required for co-ordinated (selective) operation of the various short-circuit protection devices.

The procedure to follow in using this material to make such a study can be summarized in just a few steps.

1. Secure the information needed and make the necessary short-circuit calculations to determine the different values of minimum and maximum fault current required to permit selection of co-ordinated time-current curves for the various devices involved.

2. Using the proper values of fault current from item 1, select time and current settings for the adjustable devices and check the performance of the nonadjustable devices with the objective of securing a combination that will operate in the sequence necessary to isolate a fault with minimum disturbance to unfaulted portions of the system.

3. Plot time-current curves for those devices that are to operate selectively in series to be sure that there are no potential trouble spots due to unexpected overlapping of time-current curves, or unnecessarily long operating time intervals between devices in series.

Actual plotting of the curves is desirable because rarely will all of the fault protective devices involved have the same shape time-current curves, and it is difficult to visualize mentally the relationship of the many different shapes of curves. They should be plotted on a single sheet of graph paper using a common current scale. It seems to work out best to use a scale corresponding to the currents expected at the lowest voltage level, for example, for fault current protective devices on both sides of a 2,400-480-volt transformer plot everything on a 480-volt current scale. To plot 2,400-volt device time-current curves on the 480-volt scale, first determine the desired time and current settings in the usual manner on the basis of current expected in the 2,400-volt circuit. Then plot time directly since that scale is unchanged, but multiply the 2,400-volt currents by 5 (ratio of voltages) before plotting on the 480-volt scale.

Preferably the curves should be plotted progressively as each circuit is studied starting with the device at the end of the chain (farthest from the source). This procedure will show whether or not the proposed time-current characteristic of each successive device co-ordinates with the one on its load side.

No Discussion

Aircraft Radio Interference Measurements

M. M. NEWMAN
ASSOCIATE MEMBER AIEE

R. C. SCHWANTES
ASSOCIATE MEMBER AIEE

J. R. STAHMANN
ASSOCIATE MEMBER AIEE

IT IS essential to know the wave shapes and statistical distribution of impulsive interference transients for working out adequate interference-reduction methods. Interference pulses such as are produced by corona and sparking phenomena in electric equipment present a serious problem in measurements. Conventional noise meters convert any kind of pulse to the same fundamental shock excitation wave shape determined by the characteristics of the initial circuits and thus the original character of pulses is masked. The noise meter is useful in assisting in the routine elimination of noise from various electric devices, for it informs the investigator of the progress made and it indicates when the noise is of such a character that it will no longer cause appreciable disturbance in a typical receiver. However, for interference studies covering a wide frequency range, evaluation of the noise level requires a vast number of laborious measurements with different instruments for each range of frequency and type of communications. When evaluating atmospherics or other interference that changes character with time, noise meter measurements cannot be made quickly over a wide frequency range. Thus for a fundamental study of the nature of interference, the oscillograph provides more useful information. A high-speed oscillograph is of value even in routine suppression tests, since it can show the effect of each suppressive device on the undesirable transient graphically.

Typical Impulsive Interference Transients

Impulsive radio interference transients generally consist of steep-wave-front electric impulses which shock-excite the tuned circuits of the receiver. In the case of aircraft, these pulses may be generated directly on the antenna system or may be transferred to the receiver from interfering electric equipment by coupling from the conductors of the interior electric wiring. In either case, a damped sinusoidal wave train of sufficient duration to cause an audible disturbance at the loud speaker is set up in the tuned circuits of the receiver by each pulse. The maximum amplitude of the oscillations will depend on the rate of voltage rise and amplitude of the interference pulse and, within limits, on the duration of the pulse, while the frequency and decrement depend on the constants of the receiver.

The oscillograms of Figures 1, 2, and 3 indicate a fast and slow time resolution of typical interference transients. Figure 1 shows impulses from natural atmospheric causes. These are induced surges on the antenna from radiation produced by lightning discharges. In (A) there is a continuous series of relatively high rate of rise voltage variations. This type of wave form will completely blanket communications for its duration. Some atmospherics are of the form shown in (C) containing relatively short-duration pulses or groups of pulses with spaces between. Typical single pulses on a faster time base are given in (B) and (D). The oscillograms at (E) and (F) show the atmospherics wave forms superimposed on the resultant of several strong local radio frequency carriers.

Figure 2 shows several other examples of nature-made interference. From the noise reduction point of view these differ from the atmospherics of Figure 1 in that it is possible in almost all cases to remove the source of interference. Figures 2(A)

Paper 52-241, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23-27, 1952. Manuscript submitted March 25, 1952; made available for printing May 14, 1952.

M. M. NEWMAN, R. C. SCHWANTES, and J. R. STAHMANN are with the Lightning and Transients Research Institute, Minneapolis, Minn.

The research assistance by the Office of Naval Research and the Navy Department, Bureau of Aeronautics, is gratefully acknowledged. Phases of the work have been advanced through researches with United States Air Force Wright-Patterson Communication and Navigation Laboratory, and Components and Systems Laboratory. Also co-operation of the Allen B. Du Mont Laboratories through a research grant and special cathode-ray tube construction has greatly furthered the oscillographic phases.

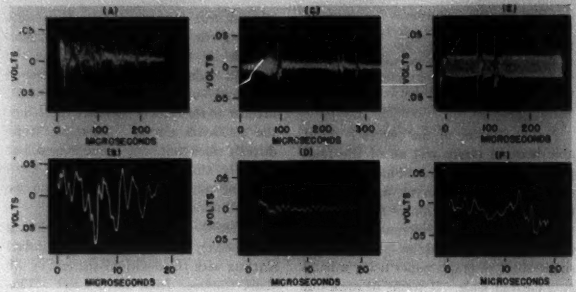


Figure 1. Typical atmospherics wave forms

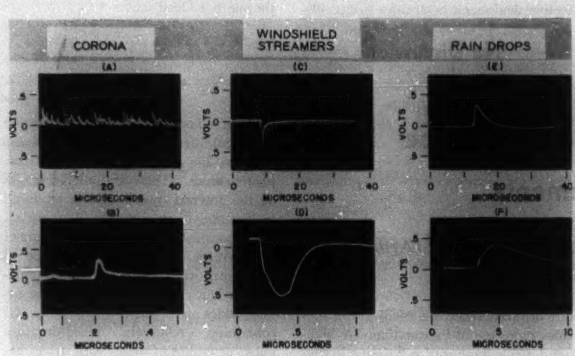


Figure 2. Typical nature-made aircraft interference

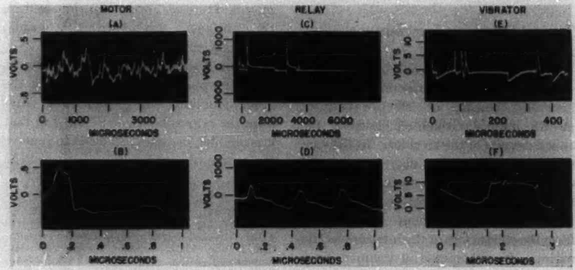


Figure 3. Electric equipment transients

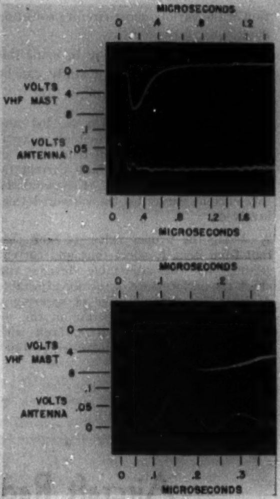


Figure 4. Direct and coupled corona pulses

and (B) show corona discharge pulses on an aircraft antenna under precipitation-static conditions. The typical individual corona pulse shown in (B) has a very steep wave front and a slower decay with a width less than 2/10 microsecond. Slow sweep oscillograms as at (A) indicate maximum repetition rates in excess of 100,000 pulses per second. Figures 2(C) and (D) show interference pulses from windshield streamers. These are caused by flashovers on the windshield surface due to a high voltage produced by frictional charging of the windshield. Repeated oscillograms have shown a maximum pulse repetition rate of 100,000 pulses per second from streamers simulated in the laboratory. Figures 2(E) and (F) show noise pulses from laboratory-produced charged raindrops striking a bare-wire aircraft antenna. These pulses have amplitudes up to 500,000 microvolts and are steep enough to cause noise at radio frequencies up to 10 megacycles.

Figure 3 shows interference from several man-made sources. Typical examples of motor commutation transients shown in Figures 3(A) and (B) vary widely with condition of brushes and present a serious problem in that the number of pulses is larger than simple commutation functions would indicate, corresponding rather to a situation of continuous chattering of brushes on the commutator surface resulting in random repetition of the type of transient generated by the single chattering relay contact. Figures 3(C) and (D) show pulses across a relay coil caused by contact chattering of a series toggle switch. Amplitudes of these pulses reach values of the order of 1,000 volts, and, since aircraft have many relays, coupled interference from this source may be considerable. The pulses in Figure 3(E) and (F) were generated by an ordinary 24-volt aircraft vibrator. In (E) is a group of pulses associated with a single contact

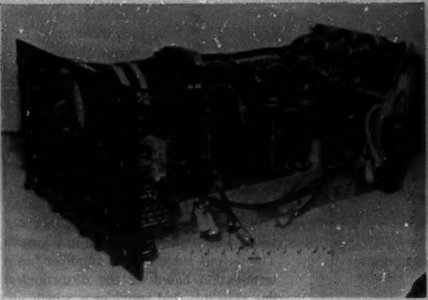


Figure 5. Basic dual-beam oscillograph

Figure 6. Oscillograph view showing 0.1–250-megacycle distributed amplifier

closing, caused by chattering of the contact.

Aircraft interference may be further classified as either direct or coupled interference. Of the aforementioned sources of aircraft radio interference one of the most common is corona which occurs directly on bare-wire antennas when the aircraft passes through thunderstorm crossfields or accummulates a charge due to frictional contact with atmospheric particles. Direct corona pulses have amplitudes of the order of 0.5 volt. However, this interference may be reduced at the source by using a polyethylene-insulated antenna which prevents corona on the antenna by a mechanism of charge deposition on the surface of the insulation sufficient to neutralize the field to such a degree that no further ionization occurs. However, during the process of variation of charge conditions, some interference potentials are induced on the antenna, which largely disappear as a steady state is reached. Interference may also be induced from corona on other parts of the plane such as the propeller and wing tips and other sharp projections from the aircraft. Figure 4 shows an oscillogram of a corona pulse on an aircraft very-high-frequency antenna mast together with the corresponding pulse coupled into a high-frequency antenna on the plane. These two pulses were recorded simultaneously on a dual-beam oscillograph. It may be seen that the coupled pulse has an amplitude of 0.07 volt which is large enough to cause radio noise.

Charged raindrops impinging on a bare-wire antenna also produce direct interference pulses, but induced pulses from this source produced when a rain drop passes near an antenna or strikes an insulated antenna probably do not have peak amplitudes and rates of rise sufficient to cause appreciable interference.

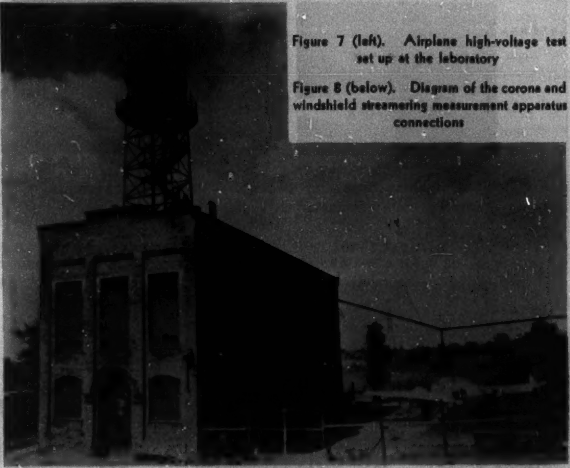
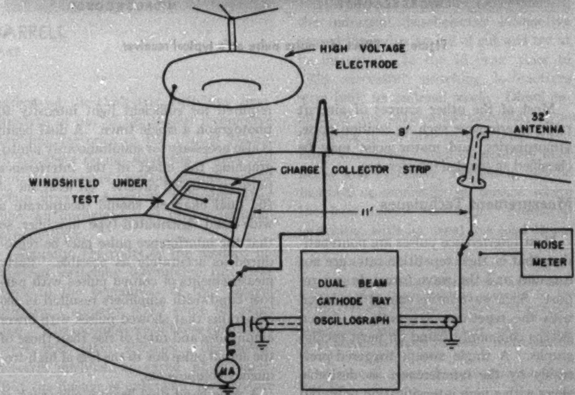


Figure 7 (left). Airplane high-voltage test set up at the laboratory

Figure 8 (below). Diagram of the corona and windshield streamering measurement apparatus connections



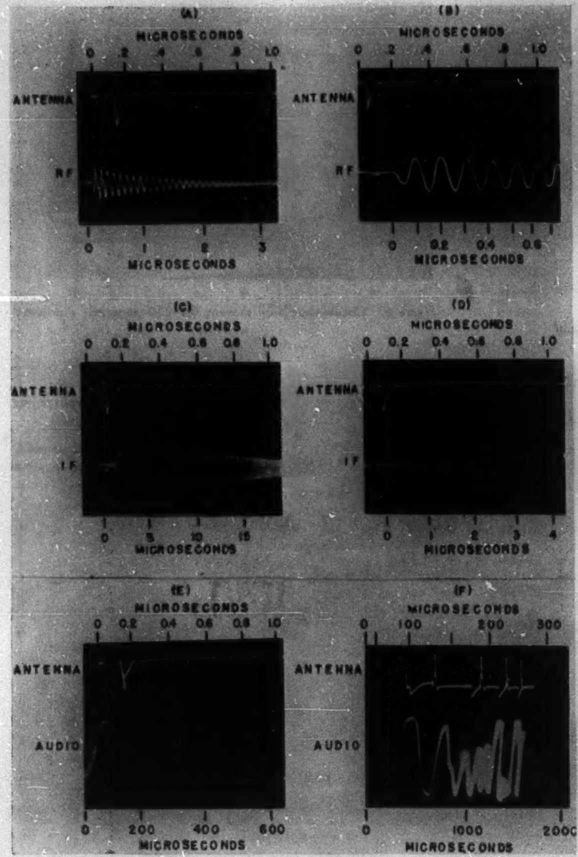


Figure 9. Effect of a noise pulse on a typical receiver

Most of the other sources of aircraft radio interference such as ignition noise, atmospherics, and motor noise may be classified as coupled interference.

Measurement Techniques

Most interference pulses are nonrepetitive, that is, their repetition rates are not constant and the wave forms do not repeat. Such wave forms cannot be viewed with the repetitive sawtooth or driven sweeps commonly found on most oscillographs. A single sweep, triggered preferably by the interference, is desirable along with a high intensification potential required for sufficient light intensity to photograph a single trace. A dual beam is also necessary for simultaneously photographing the effect of the interference pulse on coupled circuits. At least one of the dual channels should incorporate a wide-band distributed-type amplifier so that the interference pulse may be reproduced as accurately as possible. Early measurements of corona pulses with narrow bandwidth amplifiers resulted in oscillograms that showed pulses with lower amplitudes and rates of rise than those of the actual pulse due to the loss of high-frequency components.

A wide-band dual-beam oscillograph is shown in Figure 5. The unit is 14 inches high and is of relay rack width on all four sides. It has two separate channels each having all the requirements mentioned and in addition it has the driven and repetitive sweeps useful for general laboratory measurements. The cathode-ray tube is a dual beam *5XP* with an intensification potential of 25 kv. One channel has a 10-megacycle amplifier and the other has a 0.1- to 250-megacycle distributed-type wide-band amplifier. The compact wide-band amplifier is shown in Figure 6. The output stage of this amplifier has a termination based on the distortionless line principle and provides a 2-inch undistorted deflection.

To reproduce corona interference conditions, a high-voltage d-c generator is necessary to simulate the high-gradient electric fields such as may be encountered in flight. As illustrated in Figure 8, a high-gradient field was simulated by placing a high-voltage electrode above the very-high-frequency mast. This electrode was connected to a 1,000-kv d-c generator located inside the laboratory building, Figure 7, and the voltage is brought out through the wide opening shown. The dual-beam oscillograph was connected to both the very-high-frequency mast and the antenna as shown in Figure 8. The streamering interference oscillograms shown in Figure 2 were also obtained with this experimental setup.

As an illustration of the usefulness of the dual-beam cathode-ray tube, the effect of a corona pulse at various stages in a receiver is shown in Figure 9. As can be seen from the figure, the very short interference pulse is greatly lengthened by the time it reaches the audio stage as a result of shock excitation. Figures 9(D) and 9(E) expand the first part of the shock excitation wave form by increasing the sweep speed of the lower trace. Audio distortion resulting from multiple interference pulses is presented in Figure 9(F).

In studying nonrepetitive interference, it is sometimes useful to record the pulses over a long time period. The interference is photographed on a sawtooth sweep with a moving film. Such an oscillogram of motor interference is shown in Figure 10 where the sweep period is 2,000 microseconds per trace. This method is limited to relatively low-speed oscillography by the maximum film speed that can be attained to separate the traces.

While the oscillograph provides a graphical plot of the interference, the oscillograms while complete are sometimes difficult to interpret quickly. Other methods may be more useful for rapid

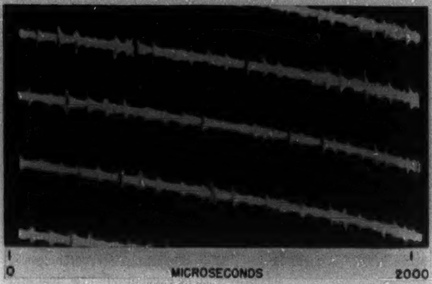


Figure 10. Motor commutation noise on a continuously moving film

noise analysis. One of these utilizes the technique of rapid electronic sampling of the outputs of a number of typical noise meters which are, at the same time, being tuned through their frequency ranges. The result is presented panoramically on one trace of a dual-beam cathode-ray tube with output amplitude plotted versus frequency. The interference pulse itself may also be presented on the other trace. This technique permits rapid coverage of a wide frequency range (0.1 to 700 megacycles in the analyzer under discussion).

Another noise-measuring device is a counter-type interference analyzer. This instrument indicates the following quantities, which appear to be the most important noise producing characteristics of impulsive interference:

1. The average number of times per second that the applied interference wave form exceeds a given rate of rise.
2. The average number of times per second that the applied interference wave form exceeds a given amplitude.
3. The average value of the interference voltage.
4. The rate of occurrence of interference pulses that coincidentally exceed the given rate of rise and amplitude in items 1 and 2 during a period when the average voltage is above a given value.

It is hoped to characterize a given noise source completely by means of several meter readings giving these quantities and with this information to predict the effects of the noise on any receiver or network over a wide frequency range.

Conclusion

For the purpose of aircraft radio noise reduction, an oscillograph capable of faithfully reproducing the interference pulses and their effects on receiving equipment provides the basic information necessary for developing effective methods of interference reduction either at the source or by circuitry. The oscillograms show the recurrence rate of the important properties of the interference from which the simplest and most effective method of noise reduction can be prescribed. For example, the low repetition rate of raindrop interference indicates that normal audio noise limiting would be effective. As a consequence of the long receiver shock excitation waveforms the high repetition rate of corona interference indicates that a noise rejection circuit located at the receiver input is necessary. The very high repetition rate and complex character of motor brush interference require noise removal at the source.

Although the oscillograph provides the basic information necessary for noise measurements, other methods as outlined either can extend the usefulness of the noise meter technique or measure the important noise characteristics directly.

No Discussion

Fifteen Years' Progress in the Design of Industrial Diesel-Electric Switchers

R. W. BARRELL
MEMBER AIEE

Synopsis: The advent of a practical line of diesel-electric industrial switching locomotives waited upon the development of high-speed automotive-type diesel engines and suitable electric transmission and control equipment. When these were made available some 15 years ago, the forerunner of the present industrial switching locomotive was designed. The intervening years have witnessed steady progress in expansion, standardization, and refinement of this line of locomotives. Today it embraces units of varying size and power adequate to meet practically every type of industrial switching duty.

THIS year marks the 15*th* consecutive year of production of standard type industrial diesel-electric switching locomotives at one of the largest locomotive building factories in the United States. While these do not have the glamor and popular appeal of their big brothers on the main lines of American railroads, they play an important part in our national productivity.

Actually, the roots of the modern industrial switching locomotive run clear back to the beginnings of the General Electric Company (GE) some 60 years ago. The locomotives of those early days were just plain electrics; the diesel-electric offshoot appeared much later. But the lineage is unbroken, for they are still fundamentally electric locomotives, only their source of power is different.

Early Diesel-Electric Locomotives

As is the case with many developments, the industrial diesel-electric locomotive passed through a period of cut and try at its inception. In the 15 years prior to 1938 successful switching locomotives were built for railroad work. Diesel engines of the high continuous-duty type operating in the 500- to 1,000-rpm speed range were used. Such locomotives were heavy and expensive, and could only be justified economically in service where the load factor was high. Industrial switching, with its usual low load factor

---

Paper 52-169, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted March 25, 1952; made available for printing April 21, 1952.

R. W. Barrell is with General Electric Company, Erie, Pa.

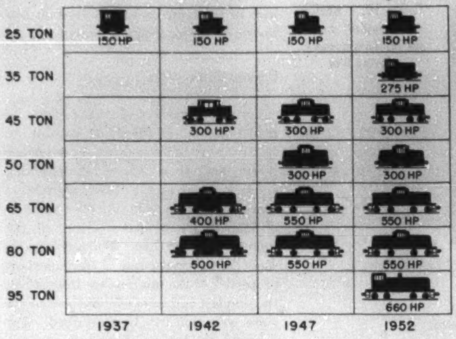


Figure 1. Development of standard diesel-electric switchers for industrial use

and often low utilization, was no place for them.

In the latter half of the 1930's engine builders achieved quantity production of low-cost high-speed automotive diesels. These quickly proved successful in the automotive field. Simultaneously, low-cost high-speed electric drives were developed for the larger sizes of these new engines.

Here were the two components for which the locomotive designers had been searching. It was only natural that they would decide to take the radical step of applying such an engine and drive to a switching locomotive. This marked the dawn of the day of the industrial diesel-electric switcher. Almost overnight the economic soundness of such a locomotive was proved, and a multimillion dollar business was born. But for this business to continue and grow, it must have as its objective customer acceptance and satisfaction. The elements entering into this may be stated in the order of their importance as follows:

1. Good performance and low operating costs.
2. Low maintenance cost.
3. Low first cost.
4. Good appearance.

The progress which has been made during the past 15 years in these elements will be discussed now.

Performance and Operating Cost

EFFICIENCY

The diesel engine is one of the most efficient prime movers known to man. As a result, the diesel-electric locomotive is decidedly more efficient than its steam brother. Since the inception of the standard industrial switcher line, a number of improvements have been made which serve to increase the overall locomotive efficiency and thus improve performance and reduce operating cost.

MOTORS

The first adaptation of high-speed traction motors to diesel-electric industrial locomotive use incorporated independent, separately designed, double-reduction gearing. The motor was originally connected to the gear unit through universal joints. When it became evident that there was a field for locomotives of this kind, the *GE-733* double-reduction motor was developed especially for this application. Subsequently this was followed by the *GE-747* motor, providing higher rated tractive efforts and higher speeds. Today, just these two types of double-reduction traction motors are used to cover the entire industrial locomotive line from 25 tons to 95 tons.

Such a wide range of locomotive sizes utilizing only two types of motors at first seems rather surprising. Nevertheless it has proved to be exceedingly practical and economical. Low speeds and maximum pulling power are provided by the use of maximum reduction gearing. On lighter weights per axle less gear reduction is used to provide a higher permissible speed. Concentration on the manufacture of these two highly specialized motors has permitted a refinement of design over the years until a high degree of standardization and interchangeability has been achieved. The pulling power of even the smallest of these economical locomotives is a revelation to those who witness their performance for the first time.

The use of double-reduction traction motors helps to account for the high efficiency of present locomotive design. A locomotive with single-reduction motors will have a motor loss of the order of 40 horsepower per motor when working near the limit of adhesion. The losses in a double-reduction motor doing the same work will be only 20 horsepower. In switching work, involving high tractive efforts but not much speed or horsepower, the gain in efficiency through the use of double-reduction motors is very real.

These motors are self-ventilated. They have been found entirely successful, even in low-speed switching service, as the double-reduction gearing amply provides

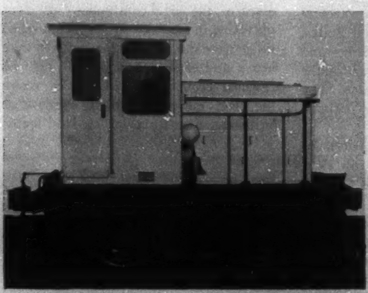


Figure 2 (left). Early diesel-electric industrial locomotive, 23 tons, 150 horsepower, forerunner of the present industrial locomotive line

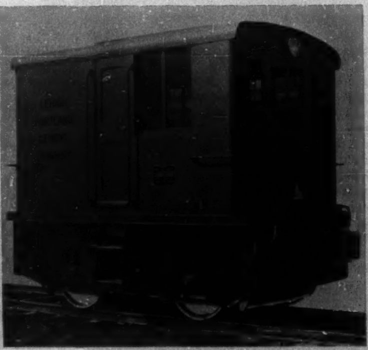


Figure 3 (right). A 25-ton 150-horsepower locomotive built in 1941, showing beginnings of the present design

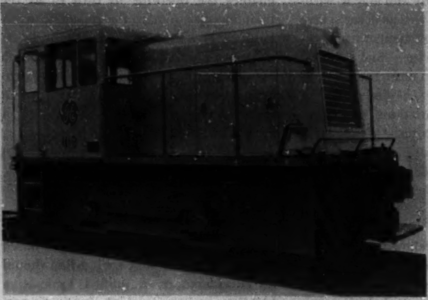


Figure 4. Modern 25-ton 150-horsepower locomotive built in 1951

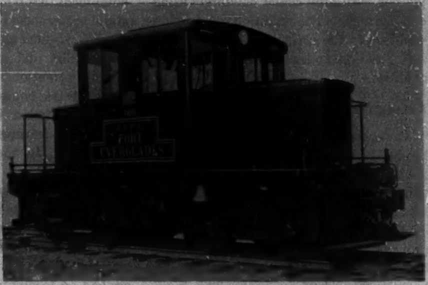


Figure 5. Earliest 45-ton 300-horsepower swivel truck locomotive with side-rod drive between axles, built in 1939

high speed for the fan mounted on the armature shaft. This design eliminates the complication of external air ducts and the power losses involved in operating separate blowers.

GENERATORS

At the beginning the existing high-speed traction-type generator was applied without change to the industrial locomotive. Some modifications were soon found desirable, the principal one being the incorporation of a cranking winding for starting the engine. Some minor mechanical changes were made, chiefly in the type of support. The modified design of generator has proved highly successful in this application, and has contributed greatly to the outstanding success of the industrial diesel-electric locomotive. Today, three types of generators cover the entire line of locomotives from 25 to 95 tons.

All these generators have simple shunt-type exciting field windings. Up to the size used on the 95-ton locomotive, they are essentially self-excited. The latter model is excited by a belt-driven exciter of the split-pole type.

CONTROL

Starting with a control system duplicating that originally developed for other traction applications, the control for the industrial locomotive line consistently has been kept to a minimum number of functions and devices. The original control provided only no power or full power. It soon became evident on early locomotives that this was unsatisfactory. If the operator wished to move the locomotive only a few feet, he might bring the engine up to full speed before anything happened, and then the locomotive would lurch forward with much greater force than desired.

A simple improvement has been made in the original control to provide battery excitation for the generator when only partial power operation is desired. Thus the engine is loaded at speeds only a little above idling and slow movements of the locomotive are accomplished easily. The generator is, however, still principally self-excited. When more power is desired the throttle handle is moved to give the engine more fuel. As the generator speeds up, the self-excitation increases and, in successive steps, the battery excitation is removed. The self-excited characteristics of the generator are such as to give practically full engine output over a wide range of current and voltage, with only a slight droop in engine revolutions per minute in the middle of the operating range. No simpler control for a diesel-electric locomotive power plant ever has been devised.

Although simple, the generator excitation and engine control now used on these locomotives up to the 95-ton size enables utilization of practically full engine horsepower over the entire locomotive speed range. In fact, the power reduction resulting from incomplete utilization of engine output is very small. It does not involve any loss of efficiency and, moreover, applies only to the maximum power obtainable at full throttle. In industrial switching service full throttle is used for only a small fraction of the work, hence the net result of this power reduction when measured in terms of work done is negligible.

From the earliest locomotive to the present models there has been an evolution in the mechanical as well as the electrical aspects of the engine control. From the beginning the engine governor has operated only at idling and top engine speed to prevent overspeed. As in an automobile or truck, the engine is controlled through the working range by giving it more or less fuel manually. On early locomotives, especially those equipped with two engines, the linkage used in the throttle control mechanism left much to be desired. In some instances, one engine would be doing all the work while the other was still running at low speed. Refinements in the design have overcome these difficulties. At the same time the new mechanism is simpler and more direct in its operation than the old. The modern locomotive line is characterized by ease of control and responsiveness to the slightest wish of the operator.

AIR BRAKES

Continuous effort has been put forth to simplify the air-brake equipment and improve brake operation. Also, much has been done to standardize air-brake equipment throughout the entire line of industrial diesel-electric locomotives. One example of the improvements introduced in the 15-year period is a better type of engineer's brake valve (self-lapping type) for straight air brakes. Another is the simplification and standardization of the combined automatic and independent brake equipment for application to this type of locomotive. Still another is the simplification and improvement of the air piping by the use of headers and pipe bends in place of numerous pipe fittings.

TRUCKS AND RUNNING GEAR

Early designs of industrial locomotives had springs directly on top of the journal boxes. As time progressed, the need for better riding qualities became increasingly apparent. The design was changed to the side equalizer type of construction, using standard freight car springs in combination with friction snubbers. At the beginning, all of the swivel-truck industrial locomotives had side-equalized suspension with drop-center equalizers using coil springs. The addition of friction snubbers in the spring nests damped the spring action and gave improved riding qualities.

Reliability

The reliability of the diesel-electric locomotive has been one of the prime factors in its wide customer acceptance and popularity. From the very first its availability was notably higher than that of its steam predecessor. Constant improvements in design and manufacture have given it an average availability of between 90 and 95 per cent. This is an important consideration for the many industrial plants operating a single locomotive. It also has a decided advantage in larger installations, where it enables the operators to meet their needs with a minimum number of locomotives.

Maintenance

The development of better materials and continual improvements in design and manufacture have done much to increase the life of the various locomotive components. Repair and replacement have been simplified greatly through improved design and better location of the parts. Particular attention has been paid throughout the design to accessibility of components. This is of great importance since a considerable amount of equipment must be installed in a limited space. Careful planning is necessary in the design in order to make the components readily accessible for servicing. As a result of close co-operation between designers and field service men a number of changes have been made from time to time to improve accessibility of parts.

Industrial locomotive service requires equipment that is sturdy and able to withstand heavy shocks as well as other severe conditions, such as high temperature, excessive dirt and moisture, and a minimum of attention. To meet these conditions, electric equipment is designed with high-temperature insulations and with mechanical features to resist shock most effectively. In the case of control devices, these shocks become such a major problem that special designs must be developed to operate successfully on diesel-electric locomotives. To illustrate the point, few devices developed for stationary installations have been successfully applied on locomotives of any kind. Instead, practically all control devices are especially designed for car or locomotive use only.

All parts of the locomotive structure are designed and constructed to resist shocks and give long, satisfactory service. The trucks and platform are fabricated from steel plates and structural shapes. The superstructure is made from sheet steel with specially formed stiffeners. Welding is used to a maximum throughout the entire structure.

Considerable progress has been made in the improvement of lubricants to prolong the life of equipment and reduce maintenance costs. The present tendency is to provide, wherever possible, long-term lubrication for roller and ball bearings so that they will not require any attention between the normal overhaul periods for the equipment in which they are used. Most traction-motor and main-generator bearings now are lubricated by this method.

The rather recent development and use of additive-type lubricating oils is doing much to prolong the life of diesel-engine parts. Thus the time between overhauls is considerably increased, with a corresponding reduction in maintenance costs.

Many of the standard components of these industrial locomotives may be handled on the unit exchange plan. This enables the operator to turn in to the manufacturer a part for repair or overhaul and receive, in return without delay, a duplicate part ready for installation on his locomotive. Such service is particularly beneficial to the industrial plant operating one or two locomotives as it eliminates the necessity of maintaining repair facilities and skilled personnel to care for the locomotives.

First Cost

Design of these locomotives was based upon the premise that such a line should be considered as a unified locomotive building project in which duplication of parts and equipment through the range of sizes could be carried out to the fullest degree. Parts common to the various sizes of locomotives were standardized so that the benefits of repetitive manufacture could be realized.

A recent example of this is found in the 45-ton locomotive, where each of the two swivel trucks is fitted with a single motor. In early designs side rods and cranks were used to transfer the power from the motored to the unmotored axle in each truck. This proved successful and was used until 1950. At that time the chain drive, which had showed itself to be very satisfactory in rigid-frame locomotives, was substituted. This change resulted in decreasing manufacturing costs, further standardization of parts, and lower maintenance expense.

The development of tools and machines for producing spot-welded stiffeners and sections has resulted in the production of a locomotive of superior quality at a low cost. A major change was made in the method of constructing the cabs when spot welding was introduced. Previously the cab had been built much like a house. A framework of structural members was arc-welded together and then the cab sheets were arc-welded to it. Now the framework is made of formed sheet steel which is spot-welded to the side, end, and roof sheets as a separate assembly. Then

Figure 6. Present design of 45-ton 300-horsepower locomotive with chain drive between axles on each truck

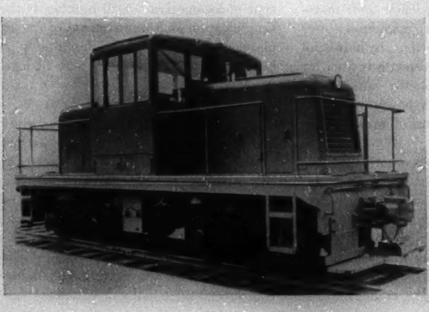


Figure 7. Streamlined design of 65-ton 400-horsepower locomotive built in 1941

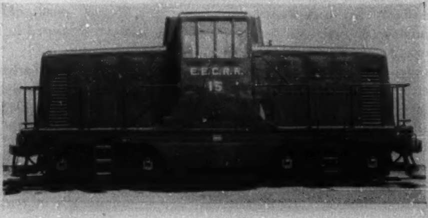


Figure 8. Present design of 65-ton 550-horsepower locomotive. The current 85-ton 550-horsepower locomotive is essentially the same

Figure 9. Present design of 95-ton 660-horsepower locomotive with single engine and operating cab at one end

the subassemblies are welded into a cab. By standardizing on the size of window openings, it is possible to use a stamping die instead of cutting them by means of a jig.

Other examples of standardization are the change in the design of the uncoupling levers, permitting them to be pressed instead of forged, and the reduction of the number of bushing sizes in the brake levers from 81 to 6.

The use of subassemblies in the manufacturing process speeds up the cycle. For instance, locomotive platforms are delivered to the assembly floor with the fuel tank and air reservoirs already installed. As much of the electric equipment as possible is assembled and wired at benches before being sent to the assembly floor. This includes the control compartment, in which the control equipment is completely assembled and wired on the bench. Operating cabs and engine cabs are produced in another section and are largely complete when brought to the locomotive assembly line.

All industrial locomotive trucks, whether of the swivel or rigid-frame type, are fabricated. The parts are burned out of steel plate and, where necessary, machined prior to their fabrication and final assembly. Wearing surfaces, such as pedestal jaws and center plates, are equipped with hardened renewable liners. When laying out parts, they are grouped on a plate so as to make maximum use of the material. In this way, small parts are cut out of what would otherwise be scrap. In some instances, the total material waste amounts to only about 3 per cent of the weight of the truck frame plus equalizer bars.

Appearance

The duplication of parts and the use of common tools had led to a similarity of appearance throughout the industrial switcher line. Furthermore, the original unattractive, box-like form has been streamlined. In most cases this was done not only for aesthetic reasons but because it paid big dividends in manufacture.

Advantage has been taken of the progress in the paint industry. The best quality of modern synthetic paints and lacquers is used to give durable finishes. A wide range of colors is available to suit the individual customer's taste. The sleek, colorful, modern industrial diesel-electric switcher truly has a definite eye appeal.

Conclusion

Close scrutiny of the current design of the industrial diesel-electric locomotive line reveals great progress in the past 15 years. Quality of performance and appearance have improved, while the elements of cost have been reduced. Rising first costs have been offset, in a measure, at least, by design simplification and by more economical use of materials. Similarly, the constantly mounting costs of labor and material for operation and maintenance have been offset to some extent by numerous design improvements resulting in increased reliability, higher efficiency, greater ease of operation, longer life of expendable parts, and simpler maintenance.

Putting all the pieces together, we obtain a picture of progress that has resulted in definite benefits to the industrial diesel-electric locomotive user. The picture is kaleidoscopic; there are many details and elements which may be viewed in different groupings and under various classifications. But no matter how it is seen, the over-all effect is well balanced and satisfying. A complete line of locomotives, adequately covering a large segment of the industrial switching and haulage field, has been developed and stabilized. This is in striking contrast with the few locomotives of odd sizes and varied designs which were in existence 15 years ago.

The result of this development program is that today American industry has the advantage of choosing from seven standard sizes of locomotives, well established as to design and in steady production. This means that any purchaser is able to select the type and size of locomotive that meets his needs most accurately and economically. At the same time he has the advantage of obtaining a modern, efficient, standardized unit, backed by years of design experience and manufacturing know-how.

No Discussion

Stabilization Templates for Servomechanisms

OTTO J. N. SMITH
MEMBER AIEE

Method

IN the process of design, the engineer must consider the change in the system transfer function due to corrective networks added either in cascade or in the feedback loop. Instead of a trial-and-error adjustment of the network parameters, templates for different compensating networks can be easily constructed which show the change in gain and phase. The template can be reduced to a single-frequency locus if the problem is one of stabilization of a feedback system, reduction of noise, or adjusting bandwidth and speed of response of a feedback system.

The templates are used with transfer function plots on co-ordinates of decibel gain versus angle. The choice of these co-ordinates for the display of the open-loop transfer function of a servomechanism is based on the fact that the addition of stabilization networks can be shown as loci. Any cascaded network will have for each frequency a specified angle shift and a specified change in gain. These can be represented by a particular length of line on this co-ordinate paper.

This method is not limited to problems of stabilization, but can be applied to many synthesis problems. The form of the compensating network usually can be determined from an inspection of pole-zero plots. A particular frequency at which the compensating network has its greatest or most significant effect is selected, and a template constructed for the gain-phase change at this particular dimensionless frequency for the family of compensating networks of the form selected. A single parameter compensator can be represented by a line template. A double parameter compensator is represented by a plane or overlay template.

The template can be moved around on the attenuation plot. The upper termination is the gain-phase point through which the compensated system will go. Depending upon the results of statistical synthesis, the gain phase of the upper termination may have widely different values for different uncompensated controlled systems.

Some very simple examples of the construction of single parameter templates follow. Applying this to stabilization, the technique adopted is to choose a simple prototype system to be controlled and compensated, select a particular stabilizing network to be used with it, and compute the locus of the change produced by this stabilizing network at the frequency of maximum closed-loop output-input ratio. This frequency is called the "critical frequency."

The template of the gain-phase shift of the stabilizing network can be used with other types of transfer functions by moving it laterally or vertically on the page. The intersection of the template and the transfer function gives the parameters of the stabilizer and the upper end of the template gives a point through which the new stabilized system will go. If the shape of the actual transfer function of the uncompensated system at high frequencies is similar to that of the prototype, then the shape of the stabilized system at the critical frequency will also be similar to that of the compensated prototype.

Integral plus Time Constant

Figure 1 shows the form of a transfer function of an integral with one time constant. In dimensionless form it is expressed by equation 1

$$G_s=\frac{\theta_0}{\theta_e}=\frac{K}{ju_1(1+ju_1)} \tag{1}$$

A family of functions and curves for different values of K are identical except that they occupy different vertical positions on the attenuation plane depending upon the value of K.

If the gain is adjusted until the prototype transfer function is tangent to the $M=1.3$ contour, then the value of K is 1.36. This is an optimum gain adjustment. If this prototype system then were cascaded with a lead network, whose low-frequency time constant is equal to the system time constant, then these two time constants would cancel and the resultant system also would appear to be an integral plus one time constant but attenuated more than the transfer function in Figure 1. If additional gain were added, and then the compensated transfer function expressed in dimensionless form, it too would have an optimum adjustment at $K=1.36$, but with the absolute frequency scale shifted toward an increased bandwidth. Figure 1 shows this optimum, or compensated, prototype system.

A single lead network can be used for two purposes. First, in case the d-c gain required of the original system is so great that it is unstable, the lead network can be used to stabilize it by introducing an extra leading phase shift at high frequencies. In case the d-c gain of the original system is low enough for stability but the speed of response or bandwidth is not great enough, it is possible to increase the high-frequency gain with a lead network and an amplifier so that the new resonant frequency is considerably higher than the original.

As a first case consider the requirement of stabilizing a controlled system whose required d-c gain is so great that the transfer function is above that in Figure 1. It cuts through the $M=1.3$ contour and is quite unstable. Let the transfer function of this be represented by equation 1 where the subscript s refers to the original or controlled system. A differentiating or lead network with a ratio of time constants of k and an additional amplifier gain of k can be used as a compensating network with a transfer function given by

$$G_c=(1+j\omega kT_2)/(1+j\omega T_2) \tag{2}$$

The transfer function of the composite system is

$$G_cG_s=\frac{K(1+j\omega kT_2)}{ju_1(1+ju_1)(1+j\omega T_2)} \tag{3}$$

For compensation, the low-frequency time constant of the stabilizing network is made to cancel the time constant of the original system

$$\begin{aligned}\omega kT_2&=u_1\\ \omega T_2&=u\end{aligned} \tag{4}$$

The frequency variable now can be shifted from u_1 to u and the system transfer function now becomes

$$G_cG_s=\frac{K}{k(j\omega T_2)(1+j\omega T_2)}=\frac{K/k}{(ju)(1+ju)} \tag{5}$$

Here the numerator coefficient has been reduced, but because of the change in the time base of the integral the gain or velocity constant is the same as before. Referring to Figure 1, it is seen that the

---

Paper 52-239, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted July 26, 1951; made available for printing May 14, 1952.

Otto J. M. Smith is with the University of California, Berkeley, Calif.

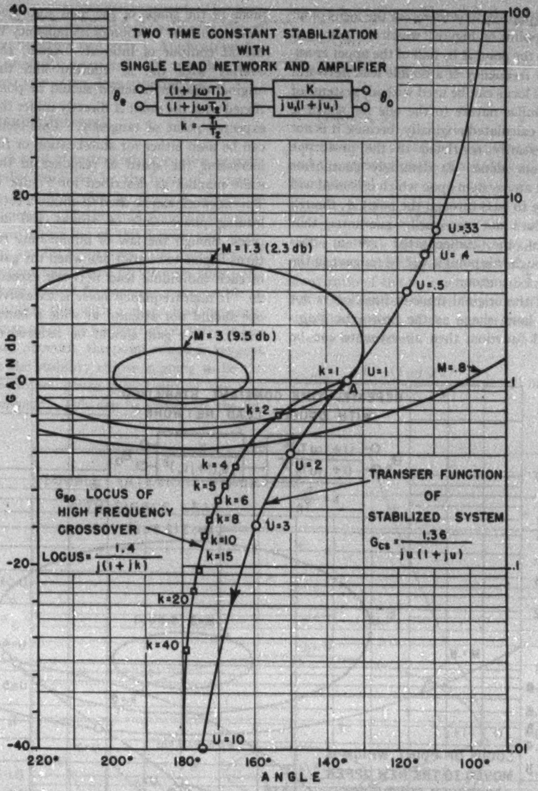


Figure 1. Locus of the change in gain and phase produced by the upper crossover frequency of a single lead network

optimum system goes through point A, a gain of 1.0 at an angle of -135 degrees. This occurs when $u=1$, or at the time constant crossover. Equation 5 is set equal to this preferred value to yield

$$(G_cG_s)_0=1.4/j(1+j)=-0.7-0.7j \qquad (6)$$

The transfer function of the compensating network at this upper crossover frequency can be obtained from equation 2 by setting $u=1$

$$G_{c0}=(1+jk)/(1+j1) \qquad (7)$$

The transfer function of the system at this upper crossover frequency can be obtained by dividing equation 6 by equation 7

$$G_{s0}=1.4/j(1+jk) \qquad (8)$$

Equation 8 is the equation for the point on the page of Figure 1 which is occupied by the original transfer function at the frequency which will eventually become the critical or resonant frequency of the stabilized system. Equation 8, therefore, describes the locus of all those points on Figure 1 which can be moved to point A by a single lead network at the upper crossover frequency. This locus is plotted; it has only one parameter, the ratio of the time constants of the original lead network. It shows that the greater the ratio of the time constants, the greater the phase shift that is available. The law of diminishing returns comes into play quickly, for after a ratio of about 10, further increases in the ratio of time constants and in the resultant gain required to construct this stabilizing network result only in about five more degrees of phase shift.

The locus is a template which is used as follows: the original transfer function is plotted on similar transparent co-ordinate paper. It is laid on top of Figure 1. The intersection of the original transfer function with the stabilizing locus yields two parameters. The k value on the stabilizing locus gives the ratio of time constants required in the lead network and the u value on the original transfer function gives the value of ωT_2 of the lead network.

With equations 2 and 4 the stabilizing system is completely specified.

In the second case where the original system is stable but the speed of response is not sufficient, the procedure is as follows: the original transfer function is laid over the template of Figure 1 and then it is moved in a vertical direction until the intersection of the original transfer function and the stabilizing locus occurs at the desired new critical frequency or new resonant frequency. The intersection gives the lead network design and the vertical shift is the additional decibel gain required.

If the original transfer function is not the same shape as the prototype stabilized function in Figure 1, then after it is stabilized its tangency with the $M=1.3$ contour, designated by A, might not occur at $-0.7-0.7j$. If the slope of the original at high frequency is greater than the slope of the prototype, A will be higher on the M contour. This is taken care of by tracing the desired M contour on the transparent overlay with the original transfer function. This overlay is then placed on Figure 1 with the point A directly under the spot on the M contour where it is expected that the compensated system will be tangent.

This template represents a pole-zero pair on the negative real axis. In the example just given, the zero was spotted on top of a pole of the original uncompensated system. In case noise is excessive, the zero should be closer to the $j\omega$ axis, or the template upper termination should be moved up and to the right on the attenuation plot.

Two-Time Constants

Another common form of transfer function is a 2-time-constant system. If the first time constant is considerably greater than the second, it can be considered as an integral so far as its effect on the region of interest for stability calculations. This is shown by the introduction of the large constant r and the approximation in which the number 1 is neglected in comparison with ωrT_b.

$$G_s=\frac{K'}{(1+j\omega T_a)(1+j\omega T_b)}$$
$$=\frac{K'}{(1+j\omega rT_b)(1+j\omega T_b)}$$
$$=\frac{K'/r}{j\omega T_b(1+j\omega T_b)} \quad (9)$$

Integral plus Repeated Time Constants

Figure 2 shows the compensation of an integral plus a repeated time constant by using a double lead network. The original transfer function family is given by

$$G_s=K/ju_1(1+ju_1)^2 \quad (10)$$

The optimally damped member of this family has K equal to one-half. The compensating network transfer function is given by

$$G_c=(1+j\omega kT_2)^2/(1+j\omega T_2)^2 \quad (11)$$

For purposes of compensation the low-frequency time constant of the compensating network is set equal to the time constant of the original system.

$$\omega kT_2=u_1$$
$$\omega T_2=u \quad (12)$$

The frequency variable also is changed so that the compensated system can be expressed in dimensionless form. It is

$$G_sG_c=\frac{K}{j\omega kT_2(1+j\omega T_2)^2}=\frac{K/k}{ju(1+ju)^2} \quad (13)$$

Figure 2 shows the transfer function of the stabilized system with the gain setting so chosen that it is tangent to the $M=1.3$ contour at A. The value of this transfer function at the point of tangency is approximately

$$G_{s0}G_{c0}=\frac{0.5}{0.45j(1+0.45j)^2}=1\underline{/-135} \text{ degrees}$$
$$=-0.7(1+j) \quad (14)$$

The value of u for this stabilized system is approximately 0.45 which means that the resonant frequency corresponds to half of the upper crossover frequency of the compensating lead network. Inserting this frequency into equation 11 yields the magnitude of the phase and gain change due to the double lead network at the new resonant frequency point. This is

$$G_{c0}=(1+0.45kj)^2/(1+0.45j)^2 \quad (15)$$

Dividing equation 14 by equation 15 gives the equation for the point occupied by the original transfer function which eventually is moved to the tangency point by the stabilizing network. This is

$$G_{s0}=\frac{-0.7(1+j)(1+0.45j)^2}{(1+0.45kj)^2} \quad (16)$$

Equation 16 is therefore the locus of all the points on Figure 2 which can be moved into the point A by half of the upper crossover frequency of a double lead network. This locus can be used with all systems of a similar nature to the one for which it was calculated originally, because it is not necessarily restricted to the prototype system alone. It definitely guarantees that any system upon which it is used will have to pass through the point A, though it does not necessarily guarantee that when the compensated system goes through this point it will be tangent to the $M=1.3$ contour.

If the original transfer function is not the same shape as the prototype stabilized function, then an estimate can be made of the shape of the final stabilized system and the location of its tangency to the M contour of interest. Again, the overlay with the M contour and the original transfer function should be positioned with the point A directly under the expected point of tangency. This locus can be used either for stabilization or for increasing the speed of response in the same manner as described for Figure 1. This network can be used to obtain a total leading phase shift of almost 180 degrees, though the law of diminishing returns begins to restrict one when the gain of each individual lead network exceeds 20. If high-frequency noise is excessive, one should not attempt as wide a bandwidth. D-c gain should be reduced or

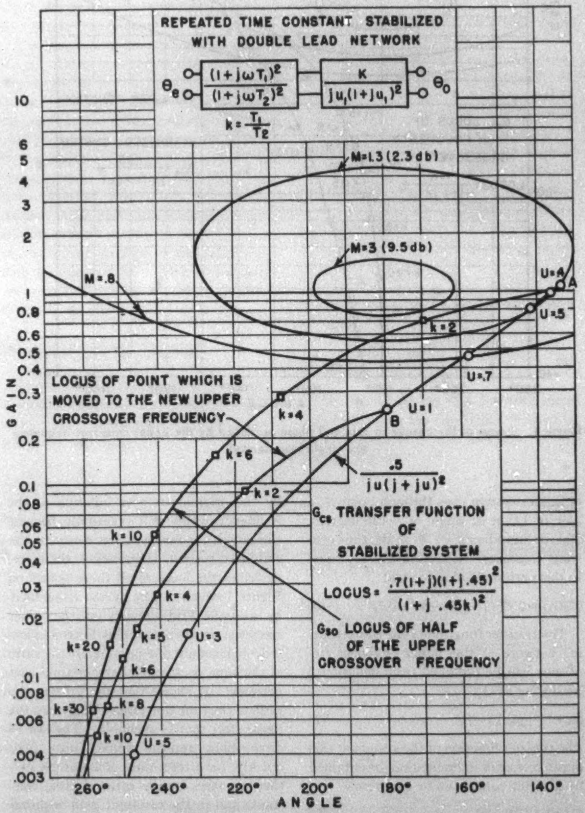


Figure 2. Locus of the change in gain and phase produced by half of the upper crossover frequency of a double lead network (change to point A). The locus terminating at point B is the change produced by the upper crossover frequency of a double lead network

complex zero stabilization used. This template represents a double pole-zero pair on the negative real axis. Applications for it are fairly restricted. A complex pole-zero pair often does a better job.

Double Integral

A kind of system which cannot be optimally stabilized by the locus of Figure 2 is the pure double integral shown in Figure 3. The original system has a constant phase lag of 180 degrees and since, for stability purposes, the compensated system must have at least an additional phase lead of 50 degrees, it is possible to stabilize this with a single lead network. Since all of the phase change is due to the lead network, the point which is tangent to the stability ellipse is going to be the maximum phase shift point of the lead network. This can be shown by the following derivation. The original system is

$$G_s=1/(ju_1)^2 \tag{17}$$

The compensating lead network is

$$G_c+\frac{(1+ju\sqrt{k})}{(1+ju/\sqrt{k})} \tag{18}$$

The maximum phase shift occurs at $u=1$, and the amount of the phase shift at this frequency is

$$\phi_m=2\tan^{-1}\sqrt{k}-(\pi/2) \tag{19}$$

For stability the phase shift should be 50 degrees and, therefore, the time constant ratio of the lead network should be 7.5. The entire transfer function of the stabilized system at the new critical frequency is therefore

$$(G_sG_c)_0=\frac{1}{(ju_1)^2}\,\frac{(1+j\sqrt{k})}{(1+j/\sqrt{k})}=-(1+j1.2) \tag{20}$$

The numerical value in equation 20 has been obtained by reading it off from Figure 3 at the tangency point. Substituting the value of k in equation 20 and solving for the critical frequency in terms of the original system, it is

$$u_{10}=1.31 \tag{21}$$

Since $u=1$ is the value of the critical frequency in terms of the compensated system, then

$$u=u_1/1.31 \tag{22}$$

The compensated transfer function in terms of original frequency variable is

$$G_sG_c=\frac{(1+ju_1 2.1)}{(ju_1)^2(1+ju_1 0.279)} \tag{23}$$

In terms of the frequency variable which is equal to 1 at the resonant frequency, equation 23 is

$$G_sG_c=\frac{0.586(1+ju\sqrt{k})}{(ju)^2(1+ju/\sqrt{k})}=\frac{0.586(1+j2.74u)}{(ju)^2(1+j.365u)} \tag{24}$$

Dividing equation 20 by equation 18 yields the value of the original transfer function at the frequency which becomes the new resonant frequency. This is also the stabilizing locus. It is

$$G_{s0}=\frac{-(1+j1.2)(1+j/\sqrt{k})}{(1+j\sqrt{k})} \tag{25}$$

It can be seen to produce at most only twice the phase shift of the locus in Figure 1 but it does so with considerably less attenuation. This stabilizing network locus can be used with any transfer function which moves across the gain angle chart in a relatively vertical direction. If the lag angle of the network exceeds 220 degrees, then no solution is possible with the single lead network, and a double lead network should be used.

Triple Integral

Figure 4 shows the stabilization of a triple integral by using a double lead network. The original system is

$$G_s=1/(ju_1)^3 \tag{26}$$

The compensating network is

$$G_c=\frac{(1+ju\sqrt{k})^2}{(1+ju/\sqrt{k})^2} \tag{27}$$

Since all of the phase shift is produced by the compensating network, the total phase shift of 140 degrees occurs at the maximum phase shift point of the double lead network and requires a ratio of time constants of 32 to 1, in accordance with the equation for each network of

$$\phi_{max}=4\tan^{-1}\sqrt{k} \quad \text{and} \quad u=1 \tag{28}$$

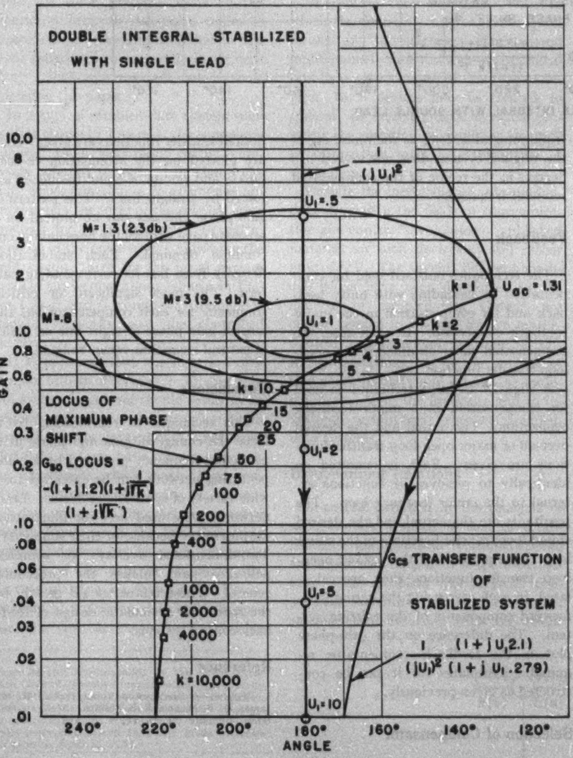


Figure 3. Locus of the change in gain and phase produced by the maximum phase-shift frequency of a single lead network

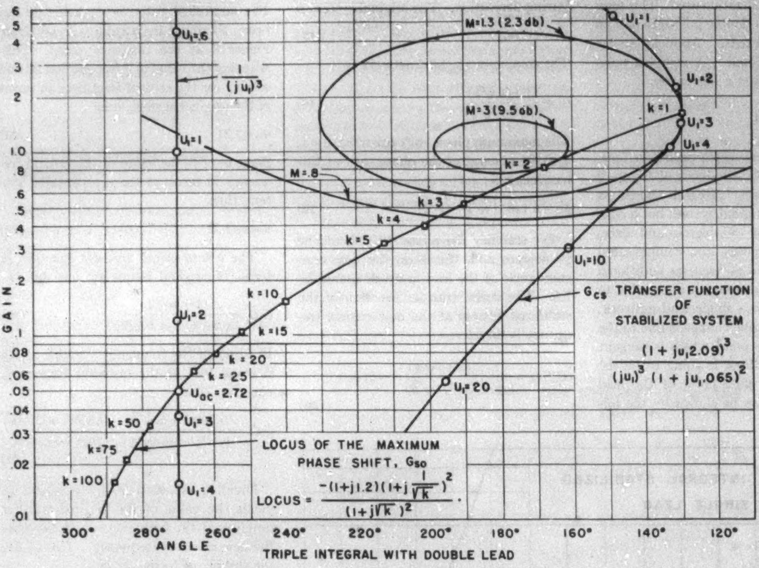


Figure 4. Locus of the change in gain and phase produced by the maximum phase-shift frequency of a double lead network

The transfer function of the stabilized system at this maximum phase shift point can be set equal to the value read off from the tangency on Figure 4, that is, $-1-j$ 1.2. This is done in equation 29

$$(G_sG_c)=\frac{1}{(ju_1)^3}\,\frac{(1+j\sqrt{k})^2}{(1+j/\sqrt{k})^2}=-(1+j1.2) \tag{29}$$

This equation can be solved for the resonant frequency in terms of the original frequency variable and it turns out to be

$$u_{10}=2.72 \tag{30}$$

The resonant frequency in terms of the variable in equation 27 is $u=1$, thus

$$u_1=2.72u \tag{31}$$

The stabilized transfer function of this system is

$$G_sG_c=\frac{1+j2.09u_1)^3}{(ju_1)^3(1+j0.065u_1)^2}$$
$$=\frac{0.05(1+j5.68u)^3}{(ju)^3(1+j0.177u)^2} \tag{32}$$

Dividing equation 29 by equation 27 at the resonant frequency yields the equation of the compensating locus.

$$G_{s0}=\frac{-(1+j1.2)(1+j/\sqrt{k})^2}{(1+j\sqrt{k})^2} \tag{33}$$

This locus shows the change in gain and phase due to the maximum phase shift point of a double lead network. It can be used on any system which resembles a triple integral as long as the phase lag of the original system does not exceed 310 degrees in the region of the compensated resonant frequency.

Feedback

The techniques outlined hold for compensation by cascading with unity feedback and for compensation in the major feedback connection. In each case, the compensated curve obtained is the open-loop transfer function.

A slightly modified technique is used for compensation in a minor feedback connection. The actual and the desired over-all or major open-loop transfer functions are given. Each is operated on identically to remove the functions external to the minor feedback loop. The resultants are the actual and the desired closed-loop transfer functions of the minor loop. The actual and the desired open-loop transfer functions then are calculated, in each case using the actual feed-forward components of the existing system. The difference on the gain-phase plot is the feedback compensator required. Templates for it can be constructed as given previously.

Selection of Compensator

The form of the compensator required can be determined if the actual and the desired transfer functions are given. Both are plotted on the attenuation co-ordinates, and points of equal frequency are joined by straight lines. This pattern of gain-phase changes can be plotted on a separate page with all the lines having one common terminal. Each stabilization network form has its corresponding pattern. The most significant or critical frequency for each compensator has the longest gain-phase straight line on scales of natural logarithm versus radians.

Conclusions

This technique of representing by a locus the change in gain and phase at a particular frequency produced by a compensating network can be extended to a wide variety of synthesis problems. Lead networks were used here for illustrative purposes only because of their simplicity. Two-dimensional overlays for complex pole-zero pairs makes the procedure general. These templates aid greatly in the rapid and systematic design of feedback control systems.

Reference

1. THEORY OF SERVOMECHANISMS, (book,) H. M. James, N. B. Nichols, R. S. Phillips. McGraw-Hill Book Company, New York, N. Y., 1947.

No Discussion

The Analysis of Sampled-Data Systems

J. R. RAGAZZINI
MEMBER AIEE

L. A. ZADEH
ASSOCIATE MEMBER AIEE

THERE is an important class of feedback control systems known as sampled-data systems or sampling servomechanisms in which the data at one or more points consist of trains of pulses or sequences of numbers. Such systems may have a variety of forms, a common example of which is shown in Figure 1. In the case illustrated, the sampling is performed on the control error by a so-called sampler which is indicated as a mechanical switch which closes momentarily every T seconds. The data at the output of such a switch consist of a train of equally spaced pulses of short duration whose envelope is the control error function. In some practical systems, the separation between successive pulses is controlled by some characteristic of the input signal and consequently is not constant. Such systems will not be considered in this paper.

In a typical sampled-data system such as that shown in Figure 1, the sampler is followed by a smoothing circuit, commonly referred to as hold or clamp circuit, whose function is to reproduce approximately the form of the original error function by an interpolation or extrapolation of the pulse train. Following the hold circuit, there are the usual components of the feedback loop, shown in Figure 1 as H and G, comprising amplifiers, shaping networks, and the controlled member.

It is apparent that the insertion of a sampler into an otherwise continuous control system in general should result in an inferior performance due to a loss of information in the control data. Yet, sampled-data systems have certain engineering advantages which make them preferable in some applications to continuous-data systems. The most important of these advantages is the fact that error sampling devices can be made extremely sensitive at the expense of bandwidth. An example of such a device is the electromechanical galvanometer and chopper bar. In this device a very sensitive though sluggish galvanometer is used to detect error and its position is sampled periodically by means of a chopper bar. The latter permits an auxiliary source of power to rotate a sizable potentiometer to a position determined by the clamped galvanometer needle. The process is carried out at uniform intervals and a sampled and clamped output is obtained for use in the continuous part of the control system. Bandwidth is lost through the sluggishness of the unloaded galvanometer, but the power gain is enormous. Similar devices for measurement of pressure errors, flow, or other phenomena can be devised along the same general lines.

In addition, there are some systems in which the data-collecting or transmission means are intermittent. Radars and multichannel time-division communication links are examples of this type of device. Such devices may be treated, in general, as sampled-data systems provided the duration of sampling is small by comparison with the settling time of the system.

Despite the increasing use of sampled-data devices in the fields of communication and control, the volume of published material on such devices is still rather limited.[1–10] The several different methods which have been developed for the analysis of sampled-data systems are closely related to the well-known mathematical techniques of solution of difference equations. It is the purpose of this paper to unify and extend the methods described in the literature and to investigate certain basic aspects of sampled-data systems.

Input-Output Relations

A central problem in the analysis of sampled-data systems is that of establishing a mathematical relation between the input and output of a specified system. This problem has received considerable attention in the literature of sampled-data systems, with the result that several different types of input-output relations have been developed, notably by Shannon,[1] Hurewicz,[2] and Linvill.[3] Shannon's relation involves the Fourier transforms of the sampled input and output; Hurewicz's relation is based on the use of so-called generating functions, which in this paper are referred to as z-transforms and which are, in fact, a disguised form of the Laplace transforms; while Linvill's relation involves directly the Laplace transforms of the input and output. The principal difference between Shannon's and Hurewicz's relations on the one hand, and that of Linvill on the other, is the fact that the former yield only the values of the output at the sampling instants, while the latter provides the expression for the output at all times, though at the cost of greater labor.

The analysis presented in this section has a dual objective: to achieve a unification of the approaches used by Shannon, Hurewicz, and Linvill; and to formulate the input-output relations for the basic types of sampled-data systems. In the next section, the problem of establishing a relation between the input and output will be approached from a significantly different point of view. Specifically, a sampled-data system will be treated as a time-variant system and its behavior will be characterized by a system function which involves both frequency and time.

The basic component of sampled-data systems is the sampler, whose output has the form of a train of narrow pulses occurring at the sampling instants 0, $\pm T$, $\pm 2T, \ldots$, where T is the sampling interval; see Figure 2. The frequency $\omega_0 = 2\pi/T$ is called the sampling frequency.

For purposes of mathematical convenience, it is expedient to treat the output pulses as impulses whose areas are equal to the values of the sampled time function at the respective sampling instants. (This is permissible provided the pulse duration is small compared with the settling time of the system and the gain of the amplifier following the sampler is multiplied by a factor equal to the time duration of the sampling pulse.) Thus, if the input and output of the sampler are denoted by $r(t)$ and $r^*(t)$ respectively, the

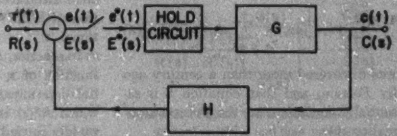


Figure 1. Typical sampled-data control system

Paper 52-161, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted March 21, 1952; made available for printing April 16, 1952.

J. R. RAGAZZINI and L. A. ZADEH are both with the Department of Electrical Engineering, Columbia University, New York, N. Y.

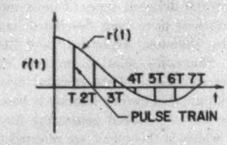


Figure 2. Pulse train at output of sampler

relation between them reads

$$r^*(t)=r(t)\delta_T(t) \tag{1}$$

where $\delta_T(t)$ represents a train of unit impulses (delta functions)

$$\delta_T(t)=\sum_{n=-\infty}^{\infty}\delta(t-nT) \tag{2}$$

Equation 1 may be written equivalently as

$$r^*(t)=\sum_{n=0}^{\infty}r(nT)\delta(t-nT) \tag{3}$$

where the negative values of n are absent by virtue of the assumption that $r(t)$ vanishes for negative values of t.

Equation 3 furnishes an explicit expression for the output of the sampler. It is more convenient, however, to deal with the Laplace transform of $r^*(t)$, which is denoted by $R^*(s)$

$$R^*(s)=\mathcal{L}\{r(t)\delta_T(t)\} \tag{4}$$

One expression for $R^*(s)$ can be obtained at once by transforming both sides of equation 3; this yields

$$R^*(s)=\sum_{n=0}^{\infty}r(nT)\epsilon^{-nTs} \tag{5}$$

An alternative expression for $R^*(s)$ can be obtained by expressing $\delta_T(t)$ in the form of a complex Fourier series

$$\delta_T(t)=\frac{1}{T}\sum_{n=-\infty}^{\infty}\epsilon^{jn\omega_0 t} \tag{6}$$

and substituting this expression in equation 1. Then, transforming the resulting series term by term there results

$$R^*(s)=\frac{1}{T}\sum_{n=-\infty}^{\infty}R(s+jn\omega_0) \tag{7}$$

where $R(s)$ is the Laplace transform of $r(t)$. It is of interest to note that the equivalence between the two expressions for $R^*(s)$

$$\frac{1}{T}\sum_{n=-\infty}^{\infty}R(s+jn\omega_0)=\sum_{n=0}^{\infty}r(nT)\epsilon^{-nTs} \tag{8}$$

was discovered more than a century ago by Poisson, and that equation 8 is essentially equivalent to the Poisson summation rule.

An inspection of either of the alternative expressions for $R^*(s)$ indicates that when s in $R^*(s)$ is replaced by $s+jm\omega_0$, where m is any integer, the resulting expression is identical with $R^*(s)$. This implies that $R^*(s)$ is a periodic function of s with period $j\omega_0$; thus

$$R^*(s+jm\omega_0)=R^*(s) \tag{9}$$

where m = any integer.

The infinite series expression for $R^*(s)$ given by equation 5 readily can be put into closed form whenever $r(t)$ is a linear combination of products of polynomials and exponential functions. For instance, when $r(t)=\epsilon^{-at}$, the right-hand member of equation 5 is a geometric series which upon summation yields

$$R^*(s)=\frac{1}{1-\epsilon^{-aT}\epsilon^{-Ts}} \tag{10}$$

When expressed in the form given by equation 5, the transform of the pulsed output, $R^*(s)$, is a function of ϵ^{sT}. This suggests that an auxiliary variable $z=\epsilon^{sT}$ be introduced and that $R^*(s)$ be written in terms of this variable. When this is done, the function $R^*(s)$, expressed as a function of z, is called the z-transform of $r(t)$. For notational convenience it is denoted by $R^*(z)$ although strictly speaking it should be written as $R^*(1/T \log z)$. With this convention, the transform given in equation 10, for example, reads

$$R^*(z)=\frac{1}{1-\epsilon^{-aT}z^{-1}} \tag{11}$$

In what follows, the symbols $R^*(s)$ and $R^*(z)$ will be used interchangeably since they represent the same quantity, namely the Laplace transform of $r(t)\delta_T(t)$. Needless to say, in cases where $r(t)$ has the form of a sequence of impulses (or numbers equal to the areas of respective impulses), the z-transform $R^*(s)$ is simply the Laplace transform of $r(t)$, and not of $r(t)\delta_T(t)$.

It will be noted that the z-transform as defined is closely related to the generating function used by Hurewicz. However, the z-transform is a more natural concept since it stems directly from the Laplace transform of the sampled time function. It is of historical interest to note that generating functions, which, as pointed out, are essentially equivalent to z-transforms, were introduced by Laplace[11] and were extensively used by him in connection with the solution of difference equations.

In practice, $R^*(z)$ is generally a rational function of z, and its inversion, that is the determination of a function $r(t)$ of which $R^*(z)$ is the z-transform, is most rapidly carried out by using a table of z-transforms such as the one compiled in Table I. (More extensive tables of closely related types of transforms may be found in references 12 and 13.) One use of this table is in finding the z-transform corresponding to the Laplace transform of a given function. Despite its brevity, Table I is adequate for most practical purposes in view of the fact that both the Laplace transforms and the z-transforms can be expanded into partial fractions each term of which can be inverted individually.

It is important to note that the inverse of a z-transform is not unique. Thus, if $F^*(z)$ is an entry in the table and $f(t)$ is its correspondent, then any function of time which coincides with $f(t)$ at the sampling instants $0, T, 2T, 3T, \ldots$, has the same z-transform as $f(t)$. To put it another way, if $G^*(z)$ is the z-transform of some function $g(t)$, then the inverse of $G^*(z)$, as found from the table, is not, in general, identical with $g(t)$, although it coincides with $g(t)$ at the sampling instants. Thus, from the z-transform of a function one can find only the values of the function at the sampling instants. In this connection, it should be noted that the value of a time function at the nth sampling instant is equal to the coefficient of z^{-n} in the power series expansion of its z-transform (regarded as a function of z^{-1}). In cases requiring numerical computations, this property of z-transforms affords an alternative, and frequently convenient, means of calculating the values of corresponding time functions at the sampling instants.

It will be helpful to summarize at this

Table I. Abbreviated Table of Laplace and z-Transforms

	Laplace Transform $F(s)$	Time Function $f(t)$	z-Transform $F^*(z)$
(1)..	1	$\delta(t)$	z^{-0}
(2)..	ϵ^{-nTs}	$\delta(t-nT)$	z^{-n}
(3)..	$\frac{1}{s}$	1	$\frac{1}{1-z^{-1}}$
(4)..	$\frac{1}{s^2}$	t	$\frac{Tz^{-1}}{(1-z^{-1})^2}$
(5)..	$\frac{1}{s+a}$	ϵ^{-at}	$\frac{1}{(1-\epsilon^{-aT}z^{-1})}$
(6)..	$\frac{a}{s(s+a)}$	$(1-\epsilon^{-at})$	$\frac{z^{-1}(1-\epsilon^{-aT})}{(1-z^{-1})(1-\epsilon^{-aT}z^{-1})}$
(7)..	$\frac{a}{s^2+a^2}$	$\sin at$	$\frac{\sin aTz^{-1}}{1-(2\cos aT)z^{-1}+z^{-2}}$
(8)..	$F(s+a)$	$\epsilon^{-at}f(t)$	$F^*(\epsilon^{-aT}z)$
(9)..	$\epsilon^{-sT}F(s)$	$f(t-T)$	$z^{-1}F^*(z)$
(10)..	$\epsilon^{as}F(s)$	$f(t+a)$	$z^{a/T}F^*(z)$
(11)..	$\frac{1}{s-\frac{1}{T}\ln a}$	$a^{t/T}$	$\frac{z}{z-a}$
(12)..	$\frac{s}{s^2+a^2}$	$\cos at$	$\frac{1-\cos aTz^{-1}}{1-(2\cos aT)z^{-1}+z^{-2}}$

stage the basic points of the foregoing discussion:

1. A sampler transforms a function $r(t)$ into a train of impulses, $r^*(t)=r(t)\delta_T(t)$, where $\delta_T(t)$ represents a train of unit impulses with period T.
2. The Laplace transform of $r^*(t)$, $R^*(s)$, is expressible in two different but equivalent forms given by equations 5 and 7.
3. $R^*(s)$ is a periodic function of s with period $j\omega_0$, where ω_0 is the sampling frequency.
4. The z-transform of $r(t)$, $R^*(z)$, is equal to $R^*(s)$ with ϵ^{sT} in $R^*(s)$ replaced by z; that is

$$R^*(z)=\mathfrak{L}\{r(t)\delta_T(t)\} \tag{12}$$

with ϵ^{sT} replaced by z.

As a preliminary to the consideration of sampled-data feedback systems, it will be helpful to establish one basic property of z-transforms. The property in question concerns the relation between the z-transforms of the output and input of the system illustrated in Figure 3. Denoting the input by $r(t)$, the output by $c(t)$, and the transfer function by $G(s)$, this relation reads

$$C^*(z)=G^*(z)R^*(z) \tag{13}$$

where $C^*(z)$ and $R^*(z)$ are the z-transforms of $c(t)$ and $r(t)$ respectively, and $G^*(z)$ is referred to as the starred transfer function. The importance of this relation derives from its similarity to the familiar relation $C(s)=G(s)R(s)$, which would obtain in the absence of samplers. This similarity makes it possible to treat z-transforms and starred transfer functions in much the same manner as the conventional Laplace transforms and transfer functions.

The proof of equation 13 is quite simple. From inspection of Figure 3, it is evident that the Laplace transform of $c(t)$ is given by

$$C(s)=G(s)R^*(s) \tag{14}$$

and correspondingly, by applying equation 7

$$C^*(s)=\frac{1}{T}\sum_{n=-\infty}^{\infty} G(s+jn\omega_0)R^*(s+jn\omega_0) \tag{15}$$

Because of the periodicity of $R^*(s)$, the following identity is noted

$$R^*(s+jn\omega_0)\equiv R^*(s) \tag{16}$$

Consequently equation 15 reduces to

$$C^*(s)=R^*(s)\left[\frac{1}{T}\sum_{n=-\infty}^{\infty} G(s+jn\omega_0)\right] \tag{17}$$

Denoting the bracketed term by $G^*(s)$

$$G^*(s)=\frac{1}{T}\sum_{n=-\infty}^{\infty} G(s+jn\omega_0) \tag{18}$$

equation 17 assumes the following form

$$C^*(s)=G^*(s)R^*(s) \tag{19}$$

which is equivalent to equation 13. It is seen that the starred transfer function $G^*(s)$, which is related to the transfer function $G(s)$ by equation 18, may be regarded as the ratio of the z-transforms of the output and input of the sampled-data system under consideration.

The mathematical essence of the foregoing discussion is the fact that a relation of the form $C(s)=G(s)R^*(s)$ implies $C^*(s)=G^*(s)R^*(s)$. This fact *per se* is very useful in the analysis of sampled-data systems involving one or more feedback loops. In the sequel, the process of passing from equation 14 to equation 19 will be referred to as the z-transformation of both sides of equation 14. The tacit understanding exists, of course, that the quantities actually subjected to the z-transformation are the time functions corresponding to the two members of equation 14.

Among the properties of the starred transfer function $G^*(s)$ there are two that are of particular importance. First, suppose that an input of the form $r(t)=\epsilon^{st}$ is applied to the system shown in Figure 3. This input is transformed by the sampler into an impulse train $r^*(t)$ which in view of equations 1 and 6 may be written as

$$r^*(t)=\frac{1}{T}\sum_{n=-\infty}^{\infty} \epsilon^{(s+jn\omega_0)t} \tag{20}$$

Operating on this expression with the transfer function $G(s)$ gives

$$c(t)=\frac{1}{T}\sum_{n=-\infty}^{\infty} G(s+jn\omega_0)\epsilon^{(s+jn\omega_0)t} \tag{21}$$

Sampling $c(t)$ and taking note of equation 18 yields after minor simplifications the expression for the sampled response of the system to $r(t)=\epsilon^{st}$; that is

$$c^*(t)=[G^*(s)\epsilon^{st}]\delta_T(t) \tag{22}$$

where $\delta_T(t)$ denotes a train of unit impulses. It is seen that the response has the form of a train of impulses whose envelope is the bracketed term in equation 22. More specifically, this means that $G^*(s)\epsilon^{st}$ represents the envelope of the response of the system to an input of the form ϵ^{st}. Consequently, it may be concluded that the starred transfer function $G^*(s)$ relates the input $r(t)$ and the envelope of the sampled output $c^*(t)$ in the same manner as the transfer function $G(s)$ relates the input and output of N.

Another important property of $G^*(s)$ concerns the impulsive response of N which is denoted by $g(t)$. Since $G(s)$ is the Laplace transform of $g(t)$, it follows at once from equations 7 and 18 that $G^*(s)$ is the z-transform of $g(t)$, that is

$$G^*(s)=\mathfrak{L}\{g(t)\delta_T(t)\} \tag{23}$$

Consequently, $G^*(s)$ may be expressed in terms of the values of $g(t)$ at the sampling instants $t_n=nT$ or, alternatively, in terms of the system function $G(s)$ via equation 18. Needless to say, $G^*(s)$ may be obtained directly from either $g(t)$ or $G(s)$ by the use of a table of z-transforms such as Table I.

It frequently happens that the system N consists of a tandem combination of two or more systems. In particular, if N consists of two networks N_1 and N_2 with respective transfer functions $G_1(s)$ and $G_2(s)$, then the transfer function of N is given by the usual relation

$$G(s)=G_1(s)G_2(s) \tag{24}$$

and correspondingly the associated starred transfer function is given by

$$G^*(s)=\frac{1}{T}\sum_{n=-\infty}^{\infty} G_1(s+jn\omega_0)G_2(s+jn\omega_0) \tag{25}$$

which in abbreviated form will be written

$$G^*(s)=G_1G_2^*(s) \tag{26}$$

The important point noted here is that the starred transfer function $G^*(s)$ of two cascaded linear systems N_1 and N_2 which are not separated by a sampler is not the product of the respective starred transfer functions $G_1(s)$ and $G_2(s)$ but rather a new transfer function given by equation 25. On the other hand, if N_1 and N_2 are separated by a sampler, as shown in Figure 4, then from equation 19 it follows at once that the z-transform of the output of N_2 is $G_1^*(s)G_2^*(s)R^*(s)$. Hence, in this case the over-all starred transfer function is

$$G^*(s)=G_1^*(s)G_2^*(s) \tag{27}$$

Consequently, it may be concluded that the over-all starred transfer function of two or more networks cascaded through samplers is equal to the product of the starred transfer functions of the individual networks.

The expressions for the z-transforms of the output of more complex structures such as those encountered in feedback

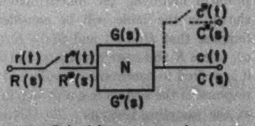


Figure 3. Pulsed linear system showing important variables and their transforms

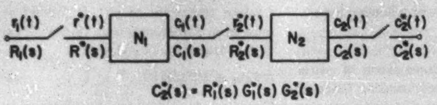


Figure 4. Cascaded linear systems separated by sampler

systems in general can be derived in a similar manner. Such expressions for several basic types of sampled-data systems are given in Table II. In this table, the first column gives the basic structure, the second gives the expression for the Laplace transform of the output, and the third gives the z-transform of the output.

It will suffice to go through the derivation of $C(s)$ and $C^*(s)$ for a typical system, say number 6 in the table. On inspection of the block diagram, it is seen that the expression for the Laplace transform of the error is

$$E(s)=R(s)-H(s)C^*(s) \tag{28}$$

where $C^*(s)$ is the Laplace transform of the input to the feedback circuit or, equivalently, the z-transform of the output of the system.

The Laplace transform of the output $C(s)$ is related to $E(s)$ by

$$C(s)=G(s)E(s) \tag{29}$$

Combining this relation with equation 28 gives

$$C(s)=G(s)R(s)-G(s)H(s)C^*(s) \tag{30}$$

Applying the z-transformation (see equation 19 and following) to both sides of this equation results in

$$C^*(s)=GR^*(s)-GH^*(s)C^*(s) \tag{31}$$

which upon solving for $C^*(s)$ yields the z-transform of the output

$$C^*(s)=\frac{GR^*(s)}{1+GH^*(s)} \tag{32}$$

as given in Table II. Finally, substituting this expression into equation 30 gives the Laplace transform of the output

$$C(s)=G(s)\left[R(s)-\frac{H(s)GR^*(s)}{1+GH^*(s)}\right] \tag{33}$$

which is the expression for $C(s)$ listed in the table.

As an illustration of the use of the expressions given in Table II, a typical problem involving the fourth structure shown in the table will be considered. Suppose that $H(s)=1$ and $G(s)=K/(s+a)$, where K and a are constants, and that it is desired to determine the values of the response of the system to a unit step input at the sampling instants. To this end, the z-transform of the output must be found. From Table II, this is

$$C^*(z)=\frac{G^*(z)R^*(z)}{1+GH^*(z)} \tag{34}$$

Referring to Table I, the z-transform of the unit step is

$$R^*(z)=\frac{1}{1-z^{-1}} \tag{35}$$

From the same table, the z-transform associated with $G(s)$ is

$$G^*(z)=\frac{K}{1-\epsilon^{-aT}z^{-1}} \tag{36}$$

Substituting these expressions in equation 34, there results after minor simplifications

$$C^*(z)=\frac{K}{(1-z^{-1})(1+K-\epsilon^{-aT}z^{-1})} \tag{37}$$

Expanding this into partial fractions

$$C^*(z)=\frac{K}{(1+K-\epsilon^{-aT})(1-z^{-1})}-\frac{K\epsilon^{-aT}}{(1+K-\epsilon^{-aT})(1+K-\epsilon^{-aT}z^{-1})} \tag{38}$$

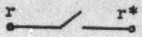
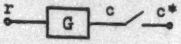
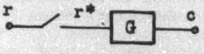
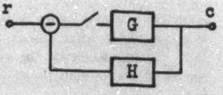
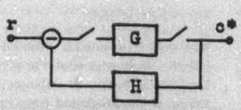
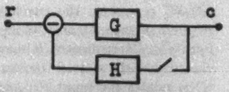
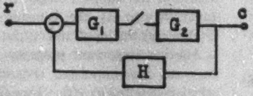
finding the respective inverse transforms from Table I and combining the results yields

$$c(t)=\frac{K}{(1+K-\epsilon^{-aT})}\left\{1-\left(\frac{\epsilon^{-aT}}{1+K}\right)^{\frac{t+T}{T}}\right\} \tag{39}$$

This function coincides with the actual output at the sampling instants and hence at the nth instant, $t_n=nT$, the value of the response to a unit step is

$$c(nT)=\frac{K}{(1+K-\epsilon^{-aT})}\left\{1-\left(\frac{\epsilon^{-aT}}{1+K}\right)^{n+1}\right\} \tag{40}$$

Table II. Output Transforms for Basic Sampled-Data Systems

System		Laplace Transform of Output $C(s)$	z-Transform of Output $C^*(z)$
(1)		$R^*(s)$	$R^*(z)$
(2)		$GR^*(s)$	$GR^*(z)$
(3)		$G(s)R^*(s)$	$G^*(z)R^*(z)$
(4)		$\frac{G(s)R^*(s)}{1+HG^*(s)}$	$\frac{G^*(z)R^*(z)}{1+HG^*(z)}$
(5)		$\frac{G^*(s)R^*(s)}{1+H^*(s)G^*(s)}$	$\frac{G^*(z)R^*(z)}{1+H^*(z)G^*(z)}$
(6)		$G(s)\left[R(s)-\frac{H(s)RG^*(s)}{1+HG^*(s)}\right]$	$\frac{RG^*(z)}{1+HG^*(z)}$
(7)		$\frac{G_2(s)RG_1^*(s)}{1+HG_1G_2^*(s)}$	$\frac{G_2^*(z)RG_1^*(z)}{1+HG_1G_2^*(z)}$

A sequence of ordinates obtained by evaluating $c(nT)$ for successive values of n yields a graph which can be used to assess the transient performance of the system. It will be noted that the system is stable for all K and a such that $\epsilon^{-aT}<1+K$. A brief discussion of the question of stability will be given in a subsequent section.

Variable Network Approach

By employing the techniques discussed in the preceding section one can obtain, in most practical cases, an explicit expression for the z-transform $C^*(s)$ and, if need be, the Laplace transform $C(s)$ of the output of a specified sampled-data system. The former can be used to find the values of the output at the sampling instants. The latter may be used, in principle, to determine the output at all times by calculating the inverse Laplace transform of $C(s)$. In practice, however, the inversion of $C(s)$ is difficult because $C(s)$ is a rational function in both s and ϵ^{sT} and no tables of inverse transforms for such functions are available.

An alternative approach which works quite well in those cases where an approximate expression for the continuous output—and not just its values at the sampling instants—is desired, is based on treating a sampling system as a periodically varying linear network. This approach is developed in the sequel, following a brief introductory discussion of the frequency analysis technique of handling time-variant systems.[14]

In using the frequency analysis technique, a linear time-variant system N is characterized by its system function $K(s;t)$, which is defined by the statement that $K(s;t)\epsilon^{st}$ represents the response of N to an exponential input ϵ^{st}. If the system function of N is known, then the response of N to an arbitrary input $r(t)$ can be obtained by superposition. More specifically, the output is given by

$$c(t)=\mathcal{L}^{-1}\{K(s;t)R(s)\} \tag{41}$$

where $\mathcal{L}^{-1}$ represents the inverse Laplace transformation and $R(s)$ is the Laplace transform of $r(t)$. The variable t in $K(s;t)$ should be treated as if it were a parameter. This implies that in evaluating the inverse Laplace transform of $K(s;t)R(s)$, one may use a standard table of Laplace transforms and treat t in $K(s;t)$ as a constant.

When N varies in time with period T, its system function $K(s;t)$ is likewise a periodic function of time with period T. Consequently, $K(s;t)$ may be expanded into a Fourier series of the form

$$K(s;t)=\sum_{n=-\infty}^{\infty} K_n(s)\epsilon^{jn\omega_0 t} \tag{42}$$

where $\omega_0=2\pi/T$ and the $K_n(s)$ represent the coefficients of the series. Thus, in the case of a periodically varying network, the problem of determination of $K(s;t)$ may be reduced to that of finding the coefficients of the Fourier series expansion of $K(s;t)$. In practice, a few terms in equation 42 usually provide an adequate approximation to $K(s;t)$, so that in many cases only two or at most three coefficients in the Fourier series expansion of $K(s;t)$ need be determined. (Note that $K_{-n}(s)$, is the conjugate of $K_n(s)$ since $K(s;t)$ is a real function of time.) Once $K(s;t)$—or, rather, an approximation to it—has been determined, the system function $K(s;t)$ can be used in the conventional manner for the purpose of obtaining the response of N to a specified input, for the investigation of the stability of N, for the calculation of the mean-square value of the response of N to a random input, for the determination of the ripple in the output, and many other purposes that are not pertinent to the present analysis.

The application of the general approach just outlined to the analysis of a sampled-data system is quite straightforward. For simplicity, the third system in Table II will be considered first. An input of the form ϵ^{st} is transformed by the sampler into a series of exponential terms which may be written as

$$\frac{1}{T}\sum_{n=-\infty}^{\infty} \epsilon^{(s+jn\omega_0)t} \tag{43}$$

The response of the network N (following the sampler) to this input is

$$\frac{1}{T}\sum_{n=-\infty}^{\infty} G(s+jn\omega_0)\epsilon^{(s+jn\omega_0)t} \tag{44}$$

where $G(s)$ is the system function (transfer function) of N. Consequently, from the definition of the system function $K(s;t)$ of the over-all system, it follows that

$$K(s;t)=\frac{1}{T}\sum_{n=-\infty}^{\infty} G(s+jn\omega_0)\epsilon^{jn\omega_0 t} \tag{45}$$

which is in effect a complex Fourier series expansion of $K(s;t)$. It is seen that the coefficient $K_n(s)$ of $\epsilon^{in\omega_0 t}$ is equal to $1/T\ G(s+jn\omega_0)$.

As a simple illustration of the use of this expression suppose that the input is a unit-step, that $G(s)=1/(s+a)$, and that ω_0 is such that $K(s;t)$ is adequately approximated by the first two terms in equation 45. In this case, $R(s)=1/s$ and

$$K(s;t)=\frac{1}{T}\left[\frac{1}{s+a}+\frac{\epsilon^{j\omega_0 t}}{s+a+j\omega_0}+\frac{\epsilon^{-j\omega_0 t}}{s+a-j\omega_0}\right] \tag{46}$$

Substituting these expressions in equation 41 and performing the inverse Laplace transformation with the help of a standard table of Laplace transforms, with t treated as a constant, one readily obtains

$$c(t)=\frac{1}{Ta}(1-\epsilon^{-aT})+\frac{2}{T(a^2+\omega_0^2)}\times \left[a\cos\omega_0 t+\omega_0\sin\omega_0 t-a\epsilon^{-aT}\right] \tag{47}$$

which is the desired expression for the output.

Turning to feedback systems, consider the fourth system in Table II. In this case, it is expedient to obtain first the expression for $e^*(t)$ corresponding to an exponential input $r(t)=\epsilon^{st}$. In view of equation 22, this is

$$e^*(t)=\frac{\epsilon^{st}}{1+GH^*(s)}\frac{1}{T}\sum_{n=-\infty}^{\infty}\epsilon^{jn\omega_0 t} \tag{48}$$

where the second factor represents $\delta_T(t)$. To deduce $K(s;t)$ from this expression, it is sufficient to find the response of the forward circuit, characterized by $G(s)$, to $e^*(t)$ and divide the result by ϵ^{st}. This yields

$$K(s;t)=\frac{1}{T[1+GH^*(s)]}\sum_{n=-\infty}^{\infty} G(s+jn\omega_0)\epsilon^{jn\omega_0 t} \tag{49}$$

which is in effect a complex Fourier series expansion of $K(s;t)$, with the coefficient of $\epsilon^{j\omega_0 t}$ being

$$K_n(s)=\frac{G(s+jn\omega_0)}{T[1+GH^*(s)]} \tag{50}$$

It will be noted that, as should be expected, at the sampling instants $t_m=mT$, $K(s;t)$ reduces to

$$K(s;mT)=\frac{G^*(s)}{1+GH^*(s)} \tag{51}$$

which will be recognized as the starred transfer function of the over-all system.

In the case under consideration, the determination of the response of the system to a given input is complicated somewhat by the fact that the denominator of $K_n(s)$ is a rational function in ϵ^{sT}, rather than in s. For purely numerical computations this is generally not objectionable. However, in analytical work it is usually necessary to approximate the term $GH^*(s)$ in equation 50 by a few terms in its expansion

$$GH^*(s)=\frac{1}{T}\sum_{n=-\infty}^{\infty} G(s+jn\omega_0)\times H(s+jn\omega_0) \tag{52}$$

In this way, one obtains a rational function approximation to $K_n(s)$.

In the final analysis, the results obtainable by the method outlined in this section are also obtainable, although less conveniently, from the expression for the Laplace transform of the output $C(s)$. The chief advantage of the system function approach is that the function $K(s;t)$ is essentially a time-varying transfer function and as such provides a clear picture of the state of the system at each instant.

Response to a Random Input

The expression for the system function obtained in the preceding section has an immediate application in connection with the important problem of determining the statistical characteristics of the response of a sampled-data system to a random input. Since a general discussion of this problem is outside the scope of the present paper, the following analysis is limited to the case where the input is a stationary time series, and it is desired to obtain the expression for the mean-square value of the output at a specified instant of time. This particular problem has considerable bearing on the design of sampled-data systems that are optimum in the sense of the minimum rms error criterion.

A general expression for the mean-square value of the output of a time-variant system is given in reference 14; it reads

$$\sigma^2(t)=\int_0^\infty |K(j\omega;t)|^2S(\omega)df,\ \omega=2\pi f \qquad (53)$$

where $S(\omega)$ is the power spectrum function of the input; $K(j\omega;t)$ is the system function with s replaced by $j\omega$; and $\sigma^2(t)$ is the mean-square value of the output at a specified instant t. To apply this equation to a sampled-data system it is only necessary to substitute the expressions for the system function of the system and the power spectrum of the input into equation 53 and carry out the necessary integration. In the majority of practical cases, the integration in question is most readily carried out by graphical means.

When the specified instant of time t coincides with a sampling instant, $t=nT$, the formula given above assumes a much simpler form. Thus, as was shown in the preceding section, for $t=nT$ the system function $K(s;t)$ becomes identical with the starred transfer function $K^*(s)$ of the over-all system. Consequently, equation 53 reduces to

$$\sigma^2=\int_0^\infty |K^*(j\omega)|^2S(\omega)df \qquad (54)$$

where $K^*(j\omega)$ is the starred transfer function with s replaced by $j\omega$; and σ^2 is the mean-square value of the output at any sampling instant.

Sometimes it is more convenient to express σ^2 in terms of the autocorrelation function[3] of the input, rather than in terms of its power spectrum. It can be shown readily that, in terms of the autocorrelation function of the input, $\psi(\tau)$, σ^2 is given by the following expression

$$\sigma^2=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}k(nT)k(mT)\psi[(n-m)T] \qquad (55)$$

where $k(t)$ is the inverse z-transform of $K^*(z)$. This expression is useful chiefly in those cases where the autocorrelation function $\psi(\tau)$ drops off rapidly with increase in τ.

The Hold System

As stated previously, the function of the hold circuit is to reconstruct approximately the original time function from the impulse train generated by the sampler. It is evident that if it were possible to realize a perfect hold circuit, a sampled-data system incorporating such a circuit would be identical with a continuous-data system. However, in general, a perfect hold circuit is not realizable because of the random nature of the time function which has to be reconstructed. Furthermore, a very important consideration in the design of hold circuits is the fact that a close approximation of the original time function requires, in general, a long time delay, which is undesirable in view of its adverse effect on the stability of the system. Consequently, the design of a hold circuit involves a compromise between the requirements of stability and over-all dynamic performance on the one hand, and on the other hand, the desirability of a close approximation to the original time function and the reduction of ripple content in the output of the system.

It should be remarked that the hold circuits commonly employed in practice are generally of the so-called clamp type, which is one of the simplest forms of hold circuits. More sophisticated types of hold circuits based on the use of polynomial interpolating functions have been described by Porter and Stoneman.[7]

A complete treatment of hold circuits cannot ignore the random nature of the time function which the hold circuit is called upon to reconstruct. Such a treatment is outside the scope of the present paper. The brief discussion which follows is concerned chiefly with some of the more basic aspects of hold circuit design.

The generation of an approximation to the original time function between two sampling instants t_n and t_{n+1}, from its values at the preceding sampling instants $t_n, t_{n-1}, t_{n-2}, \ldots$, is essentially a problem in extrapolation or prediction. An effective, though not optimum, method of generating the desired approximation is based on the consideration of the power series expansion of $r(t)$ in the typical interval from $t_n=nT$ to $t_{n+1}=(n+1)T$

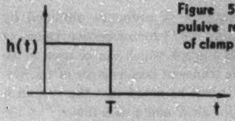


Figure 5. Impulsive response of clamp circuit

$$r(t)=r(nT)+r'(nT)(t-nT)+\frac{r''(nT)}{2}(t-nT)^2+\ldots \qquad (56)$$

where the primes indicate the derivatives of $r(t)$ at $t_n=nT$. To evaluate the coefficients of this series it is necessary to obtain the derivatives of the function $r(t)$ at the beginning of the interval in question. Since the information concerning $r(t)$ is available only at the sampling instants, these derivatives must be estimated from the sampled data. For instance, an estimate of the first derivative involving only two data pulses is given by

$$r'(nT)=\frac{1}{T}\{r[nT]-r[(n-1)T]\} \qquad (57)$$

and the second derivative is given by

$$r''(nT)=\frac{1}{T}\{r'[nT]-r'[(n-1)T]\}=\frac{1}{T^2}\{r[nT]-2r[(n-1)T]+r[(n-2)T]\} \qquad (58)$$

Thus to obtain an estimate of a derivative of $r(t)$ the minimum number of data pulses which must be considered is equal to the order of the desired derivative plus one. This implies that the higher the order, the greater the delay before a reliable estimate of that derivative can be obtained. For this reason, an attempt to utilize the higher order derivatives of $r(t)$ for purposes of extrapolation meets with serious difficulties in maintaining system stability. Generally, only the first term in equation 56 is used, resulting in what is sometimes described as a box-car or clamp circuit but which will be referred to here as a zero-order hold system. More generally, an nth order hold system is one in which the signal between successive sampling instants is approximated by an nth order polynomial.

Considering the zero-order hold system, it is evident that its impulsive response $h(t)$ must be as shown in Figure 5. By inspection, the Laplace transform of this time function is seen to be

$$H(s)=\frac{1}{s}-\frac{1}{s}\epsilon^{-Ts} \qquad (59)$$

The frequency response of this hold system is obtained by replacing the complex frequency s by $j\omega$, resulting in the following expression for the magnitude of the transfer function

$$|H(j\omega)|=\frac{2}{\omega}\left|\sin\frac{\omega T}{2}\right| \qquad (60)$$

This relation is plotted in Figure 6. It is observed that the hold system is essentially a low-pass filter which passes the low frequency spectrum of the impulse train and rejects the displaced high-frequency spectra resulting from the sampling process. An important property of the zero-order hold system is that the ripple at the output is zero if the input is a constant. In a similar way, the first order hold system has zero ripple output for an input function whose slope is a constant.

As is well known, if the signal does not contain frequencies higher than one-half the sampling frequency, perfect reproduction of the signal is obtained with an ideal low-pass filter (that is, one with unity gain and linear phase shift up to its cutoff frequency) whose cutoff frequency is equal to one-half the sampling frequency. For that matter, any low-pass network having roughly this characteristic can be used to extract most of the useful spectrum from the impulse train. It is even possible, though not advisable, to dispense with the hold circuit altogether, and rely on the low-pass characteristics of the forward circuit to perform the necessary smoothing of the sampled data.

System Design Using z-Transforms

The primary objectives in the design of closed-cycle control systems include the achievement of stability and an acceptable over-all dynamic performance. The standard technique of achieving these objectives consists in plotting the Nyquist diagram of the loop transfer function and adjusting the system parameters until the plot does not enclose the point -1.0. The margins by which such an enclosure is avoided constitute a measure of the damping of the system. Other more sophisticated techniques may be used if a better dynamic performance is desired.

In view of the fact that the starred transfer function is analogous to the conventional transfer function, the same basic technique may be applied in the case of sampled-data control systems. The function which is plotted is $GH^*(s)$ or $G^*(s)H^*(s)$ according as the denominator of the z-transform of the output is of the form $1+GH^*(s)$ or $1+G^*(s)H^*(s)$, see Table II. As usual, the complex frequency s is varied along a contour in the s-plane consisting of the imaginary axis and a semicircle enclosing the right-half of the plane. Since the starred transfer functions are periodic with period $j\omega_0$, the loci of the functions $GH^*(s)$ or $G^*(s)H^*(s)$ retrace themselves at each cycle, so that s need be varied only from $-j\omega_0/2$ to $j\omega_0/2$ in order to obtain the shape of the locus. Critical regions such as the vicinity of the origin in the s-plane are handled in the same manner as in the case of continuous-data systems.

Since the auxiliary variable z is defined as ϵ^{Ts}, it is evident that as s is assigned imaginary values over one complete cycle, z traces a unit circle in the z-plane. Thus, the loop transfer function $GH^*(s)$ can be plotted directly by expressing it in the form $GH^*(z)$ and varying z along the unit circle. To demonstrate the technique, the $GH^*(z)$ locus for the system shown in Figure 7 will be plotted. This system is seen to consist of a zero-order hold circuit and a simple linear component in the forward circuit. The feedback transmission is unity so that the loop transfer function is

$$G(s)H(s)=\frac{(1-\epsilon^{-s})}{s}\,\frac{1}{s(s+1)},\ T=1 \qquad (61)$$

Expanding this transform into partial fractions, there results

$$G(s)H(s)=(1-\epsilon^{-s})\left(\frac{1}{s^2}-\frac{1}{s}+\frac{1}{s+1}\right) \qquad (62)$$

On considering each term separately and obtaining the corresponding z-transforms from Table I, the starred loop transfer function is found to be

$$GH^*(z)=(1-z^{-1})\left[\frac{z^{-1}}{(1-z^{-1})^2}-\frac{0.632z^{-1}}{(1-z^{-1})(1-0.368z^{-1})}\right] \qquad (63)$$

which can be simplified to

$$GH^*(z)=\frac{0.264+0.368z}{(0.368-z)(1-z)} \qquad (64)$$

This function is plotted in Figure 8 where it is seen that for the constants chosen, the system is stable; but if the loop gain is increased by a factor of 1.5, the system becomes unstable.

If it were desired to improve the margin of stability or, for that matter, stabilize this system with a higher loop gain, the procedure would be to add lead networks just as in the case of a continuous-data system. The major difficulty encountered in this procedure is that the resulting starred loop transfer function is not related in a simple manner to the original function, for as was shown previously the starred transfer function of two networks in tandem is not equal to the product of respective starred transfer functions. Consequently, the insertion of a corrective network in the feedback loop requires the recalculation of the starred loop transfer function. The need for recalculation is inherent in the stabilization of sampled-data systems by the insertion of a corrective network in the feedback loop. Thus, as the art exists at the present time, the shaping procedure involves essentially a trial and error method with the plot of the Nyquist diagram for each trial set of system parameters used to assess the stability of the system.

Conclusions

The sampled-data feedback control system may be analyzed in a systematic manner by applying the z-transform method. Techniques of locus shaping similar to those commonly used in continuous systems may be applied by plotting the starred loop transfer function on the complex plane. This is done with no approximations other than those relating to the narrowness of the pulses constituting the pulse train. Once satisfactory loci are obtained by the addition of ap-

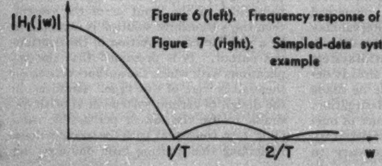
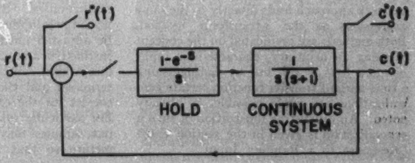


Figure 6 (left). Frequency response of clamp circuit

Figure 7 (right). Sampled-data system used in example



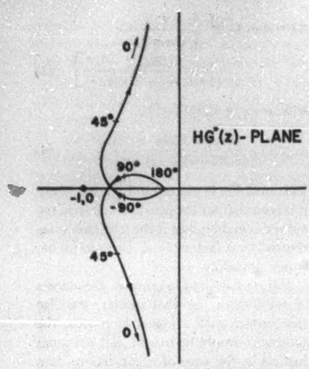


Figure 8. Plot of HG*(z) locus for system used in example

propriate networks, the transient performance of the system can be assessed by obtaining the time functions from a table of z-transforms or by expanding the z-transform of the output function into a power series which gives the ordinates at the sampling instants. The smoothness of the output can be estimated by use of the variable network analysis described in this paper.

One complication in the design of sampled-data systems is the relative difficulty of evaluating the effect of inserting corrective networks in the control loop. If this complication could be removed, locus shaping would be no more difficult than with continuous systems. Research now in progress is directed toward the devising of practical methods of locus shaping in the loop transfer function plane. Results to date indicate that the z-transform method, in conjunction with design techniques somewhat analogous to those used with conventional servomechanisms, furnishes a powerful tool for the analysis of linear sampled-data systems.

References

1. Fundamental Theory of Servomechanisms (book), **L. A. MacColl.** D. Van Nostrand Company, Inc., New York, N. Y., chapter 10, 1945.

2. Theory of Servomechanisms (book), **H. M. James, N. B. Nichols, R. S. Phillips.** McGraw-Hill Book Company, Inc., New York, N. Y., chapter 5, 1947.

3. Sampled-Data Control Systems Studied Through Comparison of Sampling with Amplitude Modulation, **W. K. Linvill.** *AIEE Transactions*, volume 70, part II, 1951, pages 1779–88.

4. Analysis of Discontinuous Servomechanisms, **F. H. Raymond.** *Annales de Telecommunication* (Paris, France), volume 4, 1949, pages 250–56, 307–14, 347–57.

5. Theory of Intermittent Regulation, **Y. Z. Tsipkin.** *Avtomatika i Telemekhanika* (Moscow, USSR), volume 10, number 3, 1949, pages 189–224.

6. The Dynamics of Automatic Controls (book), **R. C. Oldenbourg, H. Sartorius.** American Society of Mechanical Engineers New York, N. Y., 1948, chapter 5.

7. A New Approach to the Design of Pulse-Monitored Servo Systems, **A. Porter, F. Stoneman.** *Journal*, Institution of Electrical Engineers (London, England), volume 97, part II, 1950, pages 597–610.

8. A General Theory of Sampling Servo Systems, **D. F. Lawden.** *Proceedings*, Institution of Electrical Engineers (London, England), volume 98, part IV, October 1951, pages 31–36.

9. Application of Finite Difference Operators to Linear Systems, **B. M. Brown.** *Proceedings of the D.S.I.R. Conference on Automatic Control*, Butterworths Scientific Publications (London, England), 1951.

10. Analysis of a Sampling Servomechanism, **K. S. Miller, R. J. Schwarz.** *Journal of Applied Physics* (New York, N. Y.), volume 21, April 1950, pages 290–294.

11. Theorie Analytique des Probabilités, Part 1: Du Calcul des Fonctions Generatrices (book), **P. S. Laplace.** Paris, France, 1812.

12. A List of Generalized Laplace Transforms, **W. M. Stone.** *Journal of Science*, Iowa State College (Ames, Iowa), volume 22, April 1948, pages 215–225.

13. Transients in Linear Systems (book), **M. F. Gardner, J. L. Barnes.** John Wiley and Sons, Inc., New York N. Y., 1942, pages 354–356.

14. Frequency Analysis of Variable Networks, **L. A. Zadeh.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 38, March 1950, pages 291–299.

Discussion

John M. Salzer (Hughes Aircraft Company, Culver City, Calif.): The objectives of this paper are: (a) to unify the z-transform (Shannon, Hurewicz) and s-transform (Linvill) approaches to sampled-data systems; (b) to formulate input-output relations for various types of such systems; and (c) to treat such systems by considering the sampler a time-variant element.

The paper fills a clear need in this underpublished field, and it presents its topic concisely and illuminatingly. It brings together two viewpoints, (a) and (c), to bear on the same problem and shows the way toward systematization (b) of the solution.

Perhaps the most significant contribution of the paper is the variable network approach. Although the applicability of this approach was recognized before,[1] the authors' more recent investigations in the general field of time-varying systems are now made to bear directly on the problem of sampled-data systems. The importance of the generalization thus afforded should not be overlooked.

It is interesting to note that the variable network approach leads directly to the same expressions as the s-transform method. Both methods are predicated on the concept of using an arbitrary, characteristic input function [$r(t) = \epsilon^{st}$], but by distinguishing a time-variant transfer function certain generalizations are made possible, as already noted. An interesting illustration of such a generalization is given in the section titled "Response to a Random Input."

There are several points and results in this paper which bear interesting relations to some of the work done by the discussor in connection with digital-analogue systems.[2] The use of starred s-transforms, such as $R^*(s)$, in place of infinite sums, is a welcome convenience.[3] This notation facilitates the manipulations in dealing with sampled-data systems and makes the systematizations of Table II of the paper easier to comprehend.

As noted in the paper, in the case of starred transforms (that is, in the case of sampled functions) the s- and z-transforms are exactly equivalent. Thus, the use of z-transforms involves no approximation, and it offers certain conveniences. One advantage is notational, for z is easier to write than ϵ^{sT}. The other is that the infinite number of poles and zeros of the starred s-transforms are replaced by a finite set in the z-plane. Furthermore, the ambiguity at the point of infinity in the s-plane (due to essential singularity) is circumvented by the use of z-transforms.

In case the s-transform is not a starred transform, information is lost by the use of z-transform because this amounts to representing a continuous function, $c(t)$, by its samples, $c^*(t)$, just as it is done in numerical mathematics. But whereas in numerical work the sampling (or tabular) interval may be adjusted until such a representation is justified, in analyzing a given control system one is faced with a sampling rate already determined and the analysis must be made correct for the existing physical situation. For particular systems the sequence of output samples may not give a satisfactory picture so that it becomes necessary to study the behavior of the output also between sampling instants. Furthermore, and this is important, it is not always *a priori* obvious whether this is or is not the case.

Where the output behavior between samples is also of interest, the z-transform method is still applicable but must be augmented by separate investigation of the output during the sampling period. A separate solution based on initial conditions at the sampling instant may be used in analysis, but a synthesis procedure would hardly be fruitful along these lines.

The alternative solution is the s-transform approach. As noted by the authors, this method leads to transforms which are products of rational functions of s and z, and tables of corresponding transform pairs are not available. Nevertheless, the exact analysis is straightforward, even if somewhat laborious, and moreover, the frequency characteristics obtained may give a hint as to the nature of compensation needed to improve the response.

Whether the z-transform method is applicable or not in a particular case depends on the question of bandwidth. If the sampling rate is many times higher than the bandwidth of the input or of the system, then the z-transform solution is expected to be a suitable representation of the continuous output. It is presumed that the applications with which the authors concerned themselves were of this type. However, in the design of certain systems it is often desirable to use the lowest permissible sampling rate consistent with the specifications. In finding this limiting rate, one does not get a complete answer by the use of only z-transforms.

It may be noted that as far as stability is concerned either the z- or the s-transforms lead to exactly the same result, for only a divergent continuous function has a divergent sequence of samples. That this is so is demonstrated in Table II of the paper, where the corresponding denominators in both the s- and z-transform columns are identical.

To the numerous examples of sampled-data systems mentioned by the authors, the discussor wishes to add one: systems in which a digital computer is incorporated. As generally conceived today, digital computers operate on sampled data; therefore, their presence in the system requires sampling. Since the output of the digital computer is sampled also, the computer fits in between the sampler and the holding unit. If the digital computer is instructed to perform a linear difference equation on its sampled input, it can be represented by a transfer function which is rational in z. This result ties in with the methods of the paper.

For example, in the case of a digital computer equation 27 of the paper applies, because the data stay sampled through all numerical work. Thus, the transfer function of a composite digital program equals the product of the transfer functions of the component programs. The implication of this fact in system design is to be noted. Suppose that the stability of the system illustrated in Figure 8 of the paper is to be improved by a digital compensator, $W^*(s)$, inserted in either the forward or the feedback section. Since

$$WHG^*(z)=W^*(z)HG^*(z)$$

the new stability diagram is directly related to the old one so that the synthesis procedure is facilitated greatly. Of course, this is not meant to imply that digital compensation can do a better job than analogue; it only means that it is easier to see what a digital unit does. A somewhat academic example of digital compensation is worked out in chapter 4 of reference 2 of this discussion.

There is one comment concerning notation which may be found of interest: namely, it may be preferable to define z as being equal to ϵ^{-sT} rather than ϵ^{+sT}, when dealing with sampled-data systems. This is so because the latter corresponds to a time-advance operation, which has no physical meaning in a real-time application. In purely mathematical work one definition is as good as the other, and it is just unfortunate that in previous operational and transform work with difference equations the advance (ϵ^{+sT}) rather than the delay (ϵ^{-sT}) operator was given a symbol. In consequence of this choice, z naturally will appear raised to negative powers, thereby diminishing the manipulative advantage of its usage. Table I of the paper illustrates this point. Of course, multiplication of both numerator and denominator by an appropriate power of z always can eliminate the negative powers of z (as was done in line 11 of Table I), but this is an additional step.

There are further reasons for which ϵ^{-sT} should be regarded the fundamental variable. For instance, in the investigation of stability of digital programs by conformal mapping the use ϵ^{-sT} leads to much simpler rules that that of ϵ^{+sT}.

REFERENCES

1. ANALYSIS AND DESIGN OF SAMPLED DATA CONTROL SYSTEMS (thesis), W. K. Linvill. Project Whirlwind *Report R-170*, Massachusetts Institute of Technology (Cambridge, Mass.), 1949.

2. TREATMENT OF DIGITAL CONTROL SYSTEMS AND NUMERICAL PROCESSES IN THE FREQUENCY DOMAIN (thesis), John M. Salzer. Digital Computer Laboratory (microfilmed report), Massachusetts Institute of Technology (Cambridge, Mass.), 1951.

3. Compares with the notation $R(s)$ used in reference 2; for example, pages 56 and 176.

B. M. Brown (Royal Naval College, London, England): An alternative technique is available for handling the general theory and problems of the type discussed in this paper. Instead of using Laplace transforms, which imply a limitation to input and output functions which are zero for negative time, this technique assumes functions of general type, and uses as transfer functions operators which are functions of the operator $D=d/dt$. Such a function $F(D)$ is usually a fraction, in which numerator and denominator are polynomials in D. Thus if input and output, u and x respectively, are connected by the relation $x=F(D)u$, where $F(D)=P(D)/Q(D)$, then x is understood to denote the general solution of the differential equation

$$Q(D)x=P(D)u$$

Such operators have provided a classical method for solving linear differential equations. If the coefficients are constant the operators in general can be manipulated algebraically, and this type of manipulation proves to be a very powerful tool.

If a linear system has a transfer function $F(D)$, then the stability is determined by the roots of $Q(\lambda)=0$. A steady-state solution can be obtained by expanding $F(D)$ in a series of ascending powers of D and operating on u. In many cases an adequate approximation is given by the first term of this series, which always can be written down by inspection. Thus if $F(D)$ is given, the main characteristics of the response can be inferred without the labour of evaluating particular solutions to particular inputs, whether by Laplace transform or other methods.

Operators of the type $F(D)$ have parallels in the form of functions of the operators E and Δ, with a complete set of analogous properties. These operators are defined by

$$Eu(t)=u(t+T)$$

and

$$\Delta u(t)=u(t+T)-u(t)$$

where T is constant. The three basic operators are connected by the relations

$$E=1+\Delta=\epsilon^{TD}$$

the latter being obtained by Taylor's theorem. Functions of E and Δ can be used for solving difference equations and for discussing systems based on such equations. A short account of the appropriate methods is given in reference 9 of the paper.

Now operators of the types $F(D)$, $F(E)$, and $F(\Delta)$ are all linear, which is to say that, denoting such an operator by Φ

$$\Phi(u+v)=\Phi u+\Phi v$$

They also have another important property of being time-invariant, which means that if $\Phi u(t)=x(t)$, then $\Phi u(t+t_0)=x(t+t_0)$, where t_0 is any constant. Because of this, they are commutative. They can be, and are in practice, applied both to continuous functions of t and to pulse trains.

Another operator S, the sampling operator, now is introduced whose effect is to convert a continuous function into a train of pulses. It is of course equivalent to multiplication by $\delta_T(t)$ in the notation of the paper. It easily is seen that S is linear, but not time-invariant and not in general commutative with functions of D, E, or Δ.

It will be shown briefly how the various processes described in the paper can be represented operationally. Consider first the effect of sampling a function $u(t)$ and then passing it through a network with transfer function $F(D)$. The resulting function is $F(D)Su(t)$ which is in general a continuous function. If this is sampled, either physically or for the purpose of analysis, the result is the pulse train $SF(D)Su(t)$. To determine the relation between the two pulse trains, let $f(t)$ be the response to a unit impulse input to the network. Then

$$\begin{aligned}SF(D)Su(t)&=SF(D)\sum_m u(mT)\delta(t-mT)\\&=S\sum_m u(mT)F(D)\delta(t-mT)\\&=S\sum_m u(mT)f(t-mT)\\&=\sum_n\sum_m u(mT)f[(n-m)T]\delta(t-nT)\\&=\sum_n\sum_k f(kT)u[(n-k)T]\delta(t-nT),\\&\qquad\qquad\text{putting}\quad n-m=k\\&=\sum_n\left\{\sum_k f(kT)E^{-k}u(nT)\right\}\delta(t-nT)\\&=\left\{\sum_k f(kT)E^{-k}\right\}Su(t)\end{aligned}$$

The operator in brackets is a function of E. If it is denoted by $F^*(E)$ we have the operational relation

$$SF(D)S=F^*(E)S \qquad (1)$$

where

$$F^*(E)=\sum_k f(kT)E^{-k} \qquad (2)$$

Consider the special case when $F(D)=1/(D+a)$. Then $f(t)=\epsilon^{-at}$, so that

$$F^*(E)=\sum_{k=0}^{\infty}\epsilon^{-kaT}E^{-k}=\frac{1}{1-\epsilon^{-aT}E^{-1}} \qquad (3)$$

As in the paper, the general operator can be dealt with by using partial fractions and a table similar to Table I.

It is now apparent that the operator E plays a part analogous to that of the transform variable z, the association being similar to that of D and s. It will be found further that most of the equations involving starred and unstarred functions of s given in the first part of the paper can be expressed in operational form. In particular, equation 1 of this discussion corresponds to equation 13 of the paper, while the equivalent of equation 27 in the paper would be written in the operational form

$$SF_1(D)SF_2(D)S = SF_1(D)F_2^*(E)S \\ = F_1^*(E)F_2^*(E)S \qquad (4)$$

It is not easy to make a comparison of the relative merits of the two alternative approaches. So much depends on the notation with which a particular individual is familiar. It was claimed earlier that the use of operators implies greater generality, but this is perhaps of small account from a practical point of view. However, it may be an advantage to represent the operation of sampling by a special symbol.

As an example of the use of operators, consider system 4 in Table II of the paper. The circuit equation is easily seen to be

$$G(D)S[r(t)-H(D)c(t)] = c(t) \qquad (5)$$

Operating with $H(D)$ and sampling

$$HG^*(E)[Sr(t)-SH(D)c(t)] = SH(D)c(t)$$

so that

$$SH(D)c(t) = \frac{HG^*(E)}{1+HG^*(E)}Sr(t)$$

Substituting in equation 5 and simplifying gives

$$c(t) = \frac{G(D)}{1+HG^*(E)}Sr(t)$$

and

$$Sc(t) = \frac{G^*(E)}{1+HG^*(E)}Sr(t)$$

It is of interest to observe that the process of clamping can be represented by the operator $(E^{-1}\Delta/D)S$. To prove this it is only necessary to point out that a clamped function is the integral of the sequence formed by the first differences of the sampled function.

William K. Linvill (Massachusetts Institute of Technology, Cambridge, Mass.): This paper makes a concise mathematical summary of the analysis of sampled-data systems. Table II is particularly helpful and illustrates the applicability of the analysis to a wide variety of system configurations. The easy interchangeability between the z-transform method and the Laplace transform method should receive more emphasis than the paper gives. When the sampler output is considered to be a train of modulated impulses rather than a sequence of ordinates of the sampler input, the sampler has all the properties of the familiar pulse-amplitude modulator and can be treated by conventional Laplace transforms. When the whole system is treated in the frequency domain from this point of view, there is no difference between the z-transform approach and the "old-fashioned" Laplace transform approach other than a change in variable $z=\epsilon^{sT}$. When the signals at any point are discrete samples, their transforms are periodic and the $z(=\epsilon^{sT})$ variable is the convenient one to use. When the signals at a point are continuous, their transforms are aperiodic and the s variable is the convenient one to use. The attitude on the part of the engineer that the z-transform and the Laplace transform methods are equivalent and that he can change from one to the other as the situation demands is mathematically correct as well as convenient.

Freedom in changing from the z to the s domain is often particularly convenient in obtaining time responses. For example, the result of equations 46 and 47 of the paper can be seen from elementary considerations in the frequency domain. The sampler is an impulse modulator. Since $G(s)$ represents a low-pass filter, only the pure signal and the pair of lowest frequency side bands from the sampler result in significant output. The pure signal input to the filter $G(s)$ is a step of amplitude $1/T$ and the two low-frequency side bands combine to form a cosine wave: $1/T(\epsilon^{j\omega_0 t}+\epsilon^{-j\omega_0 t})=2/T \cos \omega_0 t$ for $t>0$. The output resulting from the step is $(1/aT)(1-\epsilon^{-at})$ and the output from the cosine wave is calculated from elementary transient theory to be $2/T(a^2+\omega_0^2)\,[a \cos \omega_0 t+\omega_0 \sin \omega_0 t]$ for the steady-state component and $-2a\epsilon^{-at}/T(a^2+\omega_0^2)$ for the transient. Superposing all the output components gives

$$c(t) = \frac{1}{Ta}(1-\epsilon^{-at})+\frac{2}{T(a^2+\omega_0^2)} \\ [a \cos \omega_0 t+\omega_0 \sin \omega_0 t-a\epsilon^{-at}]$$

The statement that the Laplace transformation procedure requires more labor than the other methods should be modified. To calculate the exact continuous time response is more laborious than to calculate a sequence of samples regardless of the method. The Laplace transform approach embraces both the s-domain and the z-domain pictures and from it the engineer can calculate the sampled response as easily as he can calculate it by any other method, but in addition he can get continuous signals exactly at the cost of considerable labor. Procedures have been worked out for obtaining simply an approximate continuous response and they will be described in a forthcoming paper.

J. R. Ragazzini and **L. A. Zadeh:** The authors wish to thank Dr. Salzer, Dr. Linvill, and Dr. Brown for their constructive discussions.

With regard to Dr. Salzer's comment on the variable network approach, it should be noted that the main feature of this approach is the characterization of a sampled-data system in terms of a transfer function $K(s;t)$ which involves both frequency and time. Such transfer functions do not appear in the report referred to by the discusser.[1]

In defining z as ϵ^{+sT} rather than ϵ^{-sT}, we have been motivated first by a desire to avoid conflict with the notation used by W. Hurewicz and others, and second by the fact that the alternative choice would make it inconvenient to use the only extensive table of z-transforms now available, namely, the table of so-called generalized Laplace transforms compiled by W. M. Stone. Otherwise, we are in complete agreement with Dr. Salzer's suggestion that it would be preferable to define z as being equal to ϵ^{-sT} rather than ϵ^{+sT}.

With regard to Dr. Linvill's statement to the effect that there is no difference between the z-transform and the conventional Laplace transform approaches, we believe that it would be more precise to say that the z-transform approach is a technique of analysis based on the Laplace transformation which is best suited for applications in which it is sufficient to know the values of the output at the sampling instants. By contrast, Dr. Linvill's approach leads to the expression for the Laplace transform of the continuous output, and does not yield directly the values of the output at the sampling instants. The method described in our paper achieves a unification of these approaches in the sense that it furnishes a systematic procedure for determining both the Laplace and z-transforms of the output, as illustrated in Table II. The connecting link between the two approaches is contained in the statement that a relation of the form $C(s)=G(s)R^*(s)$, where $C(s)$ and $G(s)$ are ordinary Laplace transforms and $R^*(s)$ is a starred transform (that is, z-transform), implies the relation $C^*(s)=G^*(s)R^*(s)$.

As we have pointed out in the paper, the expression for the output obtained by the use of the variable network approach also may be obtained from the expression for the Laplace transform of the output, and, in simple cases, the same results may be derived from elementary considerations in the manner indicated in Dr. Linvill's discussion. The main advantage of the variable network approach is that it yields the expression for a time-varying transfer function $K(s;t)$ which constitutes a much more explicit and flexible means of characterizing a system than the expression for the Laplace transform, $C(s)$ of the response to some particular input. Furthermore, $K(s;t)$ is, in general, more convenient to work with than $C(s)$. For example, one can readily express the mean-square value of the response to a random input in terms of $K(s;t)$, but not in terms of $C(s)$.

With reference to Dr. Brown's statement that the use of Laplace transforms implies limitation to input and output functions which vanish for negative time, it should be noted that, when such is not the case, it is merely necessary to employ the Fourier or bilateral Laplace transforms in place of the unilateral Laplace transforms. Thus, the applicability of the methods described in our paper is not restricted to input functions which vanish for negative t.

The operational approach presented by Dr. Brown is related to the z-transform approach in much the same manner as Heaviside's operational calculus is related to the conventional Laplace transformation. In particular, the operator E corresponds to z, the relation $SF(D)S=F^*(E)S$ is equivalent to the relation $C^*(s)=G^*(s)R^*(s)$, and the operation with a sampling operator S corresponds to the z-transformation. It is of interest to note that the operators employed by Dr. Brown may be regarded as special forms of so-called time-dependent Heaviside operators.[2]

A useful feature of the operational approach is that it places in direct evidence the operations performed on the operand. However, working with operators is more difficult than with z-transforms, since the latter require only purely algebraic manipulations. This is indeed the chief advantage of the z-transform method.

REFERENCES

1. See reference 1 of John M. Salzer's discussion.
2. TIME-DEPENDENT HEAVISIDE OPERATORS, **L. A. Zadeh.** *Journal of Mathematics and Physics* (Cambridge, Mass.), volume 30, 1951, pages 73–78.

Electrification of Holland's Railway System

A. H. CANDEE
ASSOCIATE MEMBER AIEE

Synopsis: The railways of the Netherlands were almost totally destroyed at the close of World War II. This provided an opportunity to change from the former system of electrification at a minimum cost if desired. Reaffirming the earlier decision to electrify with the 1,500-volt d-c system, remarkable rehabilitation has taken place and electrification is being extended rapidly. Thus, 7 years after the cessation of hostilities, the Netherlands Railways may be considered one of the most progressive of Europe, taking full advantage of the economies and improved performance accruing from the use of this simple and intensely practical system of electric propulsion.

THE adoption of trolley electric propulsion by a railroad usually is based upon the economies effected thereby or is for the purpose of performing a service which is best done by this type of power. Here in North America, it is difficult for trolley electrification to show great saving over diesel electrification and, as a result, the railroads have adopted diesel electric motive power chiefly because of the ease of financing and introducing this type of locomotive. In Europe and other parts of the world, where liquid fuels are not as cheap nor readily available, trolley electrification usually effects large enough operating savings over diesel electric motive power to effect its adoption in preference to the latter.

Central station electric energy is an economical and very versatile servant of the human race. In fact, this versatility has imposed a handicap on the progress of railroad electrification because there are so many systems by which this power may be supplied and utilized for railroad propulsion purposes and such a wide divergence of opinion as to which is the best that there has been little attempt to standardize or promote interchangeability of equipment. This has resulted in a galaxy of dissimilar trolley electrifications scattered throughout the world, nearly all of them successful in a way but reflecting a lack of co-ordination among the proponents of this type of propulsion. Once having chosen a system, a railroad sometimes finds it expensive to change to another.

The Netherlands Railways

One railroad which chose their system of electrification more than 25 years ago and which has had an opportunity in recent years to review and change their earlier decision, if desired, is the Netherlands Railways. Trolley electrification of this railroad dates from the year 1908 when the line between Scheveningen and Rotterdam (approximately 15 miles) was operated from a 10,000-volt a-c 25-cycle overhead trolley line. In 1924 it was decided that a change to the 1,500-volt d-c system would reduce their overhead line troubles and also their traction motor problems. By 1927 the original electrification had been converted and electric services had been extended northward from Rotterdam to Amsterdam and Ijmuiden. Up to the year 1940, when the Netherlands was invaded by the Germans, the network of electrified lines had been expanded gradually.

War Damage to and Recovery of the Railroads

No industry in the Netherlands suffered as much war damage as did the railway system. When British paratroopers were dropped on the north side of the Rhine near Arnhem on September 17, 1944, a command was broadcast from the Dutch government in exile for the railroad workers of Holland to quit work and go into hiding, thus to hamper the movement of German troops and supplies in this whole area. The order was obeyed, but the 8,000 paratroopers had dropped into unsuspected German armor and strength, with the result that the advance of the Allies in this area was stemmed and the Netherlands was not freed until May 5, 1945. The Germans, then, had nearly 8 months in which to plan and carry out a vindictive and methodical wrecking of the railway system, during which time the 30,000 loyal railway workers and their families had to stay in hiding and the whole populace suffered greatly because food and fuel could not be transported. When the country finally was freed, there was little left of the railways, for practically everything had been removed to Germany or thoroughly wrecked. Rolling stock was taken, burned, or destroyed; shops were looted completely—not even a wrench or screw driver was left; copper wire and electric substation equipments were carried away; bridges were wrecked; and tracks were torn up. When the Germans capitulated, the status was as shown in Table I.

Table I

Equipment	Per Cent
Destroyed or stolen	
Locomotives	84
Passenger cars	94
Freight cars	98
Electric trains	99
Diesel electric trains	100
Destroyed	
Stations	18
Signal towers	65
Railbed	62
Bridges	78

However, 7 years later (June 1952) the Netherlands Railways is again a smoothly operating, progressive railroad, with an on-time record that would be the envy of any American railroad. Trolley electric motors cars and streamlined trains have been recovered from Germany and rehabilitated or have been salvaged from the wreckage left in the country; diesel electric motor cars and trains have been recovered and put into operating condition again; steam locomotives have been returned or used ones bought from the Allied Army surplus; trolley electric locomotives have been purchased or rented; bridges have been repaired, signal equipment reconditioned, and stations rebuilt; and many other very surprising reconstructions have been accomplished. Work is still in active progress—busy track gangs may be seen on nearly every line; grade separations are being made; bridge crews are still hard at work completing repairs; shops are being modernized; new locomotives and trains are being built; electric trolley and power supply extensions are under way; and war-damaged cars and trains are still being repaired to expand the rolling stock inventory.

Re-selection of the 1,500-Volt D-C System

After the destruction of the railroad by the Germans in 1944–45, an excellent opportunity was afforded to start all over

---

Paper 52-172, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted March 26, 1952; made available for printing April 22, 1952.

A. H. Candee is with Westinghouse Electric Corporation, East Pittsburgh, Pa.

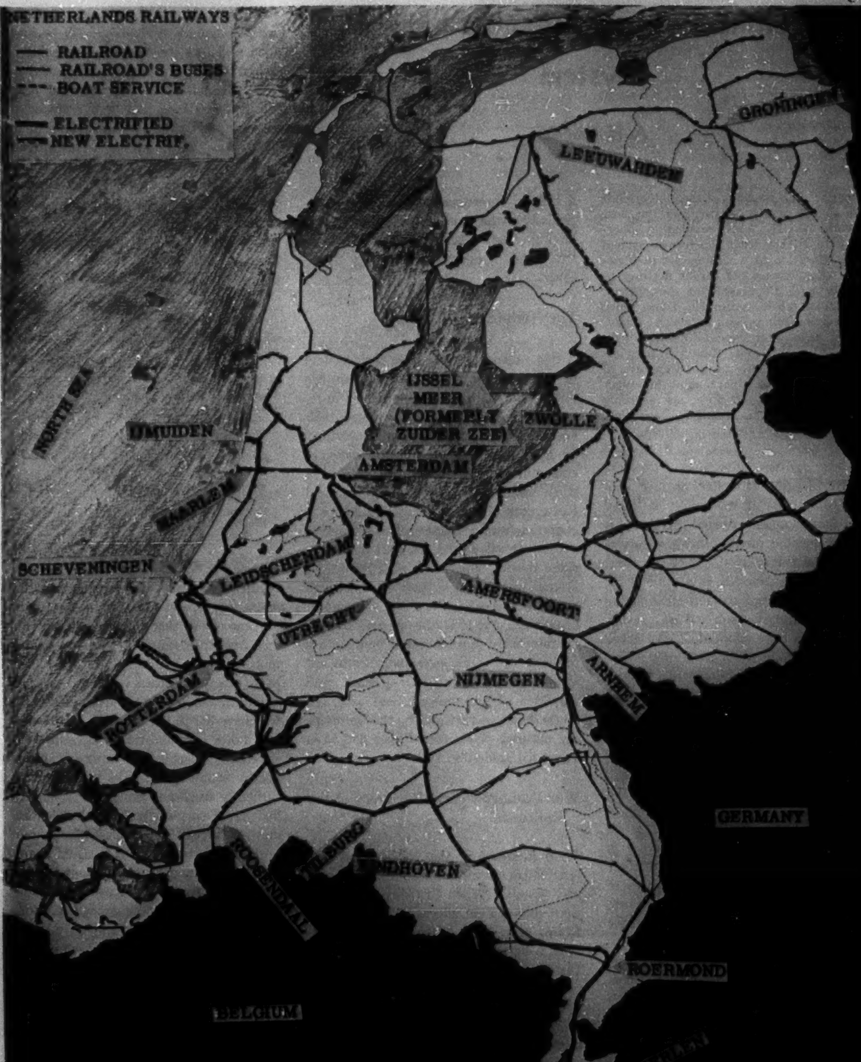

NETHERLANDS RAILWAYS
RAILROAD
RAILROAD'S BUSES
BOAT SERVICE
ELECTRIFIED
NEW ELECTRIF.
GRONINGEN
LEEUWARDEN
NORTH SEA
IJMUIDEN
IJSSEL MEER (FORMERLY ZUIDER ZEE)
ZWOLLE
AMSTERDAM
HAARLEM
SCHEVENINGEN
LEIDSCHENDAM
AMERSFOORT
UTRECHT
NIJMEGEN
ARNHEM
ROTTERDAM
GERMANY
ROOSENDAAL
TILBURG
EINDHOVEN
ROERMOND
BELGIUM
HEERLEN
MAASTRICHT

Figure 1. Map of the Netherlands

Figure 2. Trolley electric multiple-unit train

with a different power supply system if desired. A thorough study, however, confirmed the earlier decision to use 1,500-volt d-c power for the trolley lines, this being based on the density of the traffic, the prevalence of large cities and towns in relatively close proximity, and the previously demonstrated low operating costs of the 1,500-volt system. Moreover, such a medium voltage d-c system makes the operation of trolley electric motor car trains very practical at a nominal cost for equipment and repairs. Because of the predominance of these trains in the Netherlands, this factor was of considerable weight.

Since economy is of some importance in the electrification of any railroad, the selection of a system involves some definite objectives:

1. To convert from commercial power systems to a type of electric energy suitable for distribution to moving vehicles, using the simplest equipment and with the maximum efficiency.
2. To minimize the first cost of the distribution structure and equipment. The use of a voltage as high as practicable to minimize the amount of copper needed and to require the fewest number of substations assists in keeping installation costs low.
3. To minimize the cost of parasitic distribution equipment.
4. To use equipment both in the distribution system and on the vehicles which will have the lowest possible maintenance costs, because such costs are continuing over the years. Low repair costs also reflect a high degree of reliability.

The 1,500-volt d-c distribution, as decided upon by the Netherlands Railways, encompasses all of the features of an ideal system with the exception that the overhead trolley construction is somewhat heavier than might be desired and the substation spacing is not as great as for higher voltage systems. These objections were accepted, however, in order to obtain the other indisputable advantages which, for the concentrated traffic found on this railroad, far outweight the objections. The rectifiers used to convert from the 3-phase 50-cycle commercial power system to the 1,500-volt d-c system are simple, reliable, and efficient; there are no parasitic devices necessary, such as often are used with a-c systems to reduce inductive disturbances; and the substation equipment, the traction motors, and the control devices of the rolling stock have long been proved reliable and of low maintenance costs. Actually, the structures supporting the overhead wires are very little, if any, more expensive than for a high-voltage system and the weight of the wires is not excessively greater, so that the prime increase in distribution costs lies in the reduced spacing of substations. However, once installed, this factor is of reduced importance.

The Netherlands

In general, the Netherlands is flat and few important grades are encountered; see Figure 1. In the western and northern parts, much of the land is polder land (areas surrounded by dykes and usually lying below sea level), while in the east the ground is somewhat higher and is rolling. In the southeast, such as in that neck which protrudes southward between Germany and Belgium from Roermond to Maastricht and Heerlen, there are pronounced ridges and valleys. The lowest elevation of the country is approximately 20 feet below sea level and the highest point is in the extreme southeast near Heerlen where there is an elevation of 321 meters (1,050 feet). Sand is encountered almost everywhere and clay is common. The principal cities lie in the western part of the country, but Holland is actually an industrial nation and practically every town has its shops and factories.

Traffic

Internal commerce of the Netherlands depends upon three basic transport systems: the railroad, the waterways, and the highways. The railroad has far from a monopoly in the transport of freight because of a national system of canals. Since a large part of the country is built on the deltas of the Rhine, the Maas (Meuse) and the Sheldt, and since these rivers divide and subdivide into a large number of natural and artificial waterways, it followed that the early inland commerce was mostly water-borne. Add to this the vast system of canals and ditches necessary for the control of water levels throughout those areas reclaimed from the sea (which could be expanded easily into further commercial waterways) and the result is a pattern of water routes that extends even to the remote and elevated portions of the country. Actually there are about 4,800 miles of such navigable waterways—nearly $2^{1}/_{2}$ times the railway mileage—some for small barges but many for larger vessels. About 340 miles of canals are for the largest vessels (2,500 tons or larger), 1,000 miles will allow the passage of coasters (vessels plying the North Sea coastal areas), and 1,500 miles are suitable for barges up to 450 United States tons capacity.

The bulk freight of the Netherlands generally is moved by diesel engine driven canal barges—coal, coke, potatoes, vegetables, lumber, grain, peat, sand and gravel, brick and tile, and so forth. Highway vehicles as yet have not become a serious competitor for the railroad or canals, but must be given careful consideration in future planning by the railroad. In addition, there is considerable traffic in transit from the seaports of the Netherlands to inland destinations in foreign countries. Here, again, the barge is a potent competitor, plying the Rhine between Rotterdam and Switzerland or the Maas to points in Belgium. The division of the internal freight traffic is approximately as shown in Table II.

The Netherlands Railways does a very brisk passenger business, Figure 2. The railway system handles approximately 160,000,000 passengers per year for 3,900,000,000 passenger miles and requires 24,000,000 passenger train miles of operation. At the same time, the train miles in freight service total approximately 9,300,000.

Figure 3 (left). Oerlikon 1A-Bo-A1 trolley electric locomotive

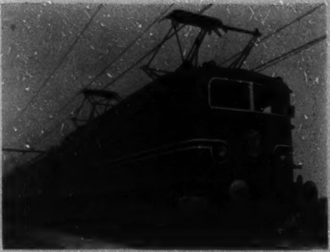


Figure 4 (right). Alsthom Bo-Bo trolley electric locomotive

Passenger services account for approximately 58 per cent of the total revenue and freight service for but 38 per cent, the balance being made up by diverse services such as mail and baggage. This indicates a somewhat different type of operation than is generally found in America, where freight traffic nearly always predominates. Holland's canal system and the paucity of automobiles undoubtedly accounts for this. On the other hand, where American railroads consider that passenger services are nonprofitable, the Netherlands Railways has an operating ratio of around 74 which is extremely favorable.

Railroad Operation

From the figures just given, it is evident that the character of the Netherlands Railways differs from that of the normal American railroad. Emphasis is placed upon passenger train services, and schedules must be so arranged that little interference is occasioned by the operation of freight trains. For this reason and because the distances to be traversed are, in general, rather short, most of the freight trains are moved at night and are off the main lines at the time that the heavy passenger traffic peaks occur. The result of this, and also of the rigid policing of passenger train delays, is that the railroad has an on-time record which should be the envy of nearly any of the railroads of the American continent. The special significance of such on-time operations is that the traveling public can rely upon ample time for making connections at the numerous interchange points even though such connections allow but a few minutes for changing from one train to another. It also allows trains to be assigned definitely to particular station tracks and interchanges usually can be made merely by walking across the island platform from one train to the other. The net outcome is that travel by railroad is a convenient, comfortable, and practical way to get from one place to another within the Netherlands.

Table II. Freight Traffic within Holland

	Per Cent
Tons of freight	
By railroad	15
By water	39
By highway	46
Ton-miles of freight	
By railroad	32
By water	53
By highway	15

The Netherlands Railways owns and operates a comprehensive network of lines extending to all parts of the country and serving practically every community in the nation. The heaviest traffic is concentrated in the western central part of this area. The Annual Report for the year 1950 showed 1,988 miles of railroad in operation, 1,055 miles of which are of single track and 933 miles of double track. Of these, 507 miles of the double track and 48 miles of the single track are electrified and work has been progressing rapidly in the extension of this system. It is expected that a total of approximately 787 miles will be electrified by the end of 1953. In the end, approximately 42 per cent of the mileage will be trolley electrified, this trackage carrying about 85 per cent of the passenger miles and 80 per cent of the freight ton miles. The ultimate use of diesel electric motor cars or locomotives has been planned for all other lines and for switching services, thus retiring all steam motive power. Right-hand running is standard, the same as in America.

The Roadway

Almost all of the main line rail is 46 kilograms per meter (92.5 pounds per yard), but some heavier rails of 63 kilograms per meter (127 pounds per yard) are on trial. Rails are of the T-head shape with flat bottoms, those for the main lines being clamped to special tie castings whch, in turn, are held in place on treated ties by screw fastenings. For yards and sidings, the rails are set on tie plates and spiked down. Rail joints are set opposite each other, generally located between ties, which is contrary to the general American practice of staggering the rail joints. Each of the cast tie plates is finished for a 1 to 20 inward cant of the rail. Rails are laid on treated ties and ballast is generally of gravel since cut stone is relatively expensive. In view of the unstable character of the soil under the roadbed, the railroad's engineers try to limit the weight per axle to 16 metric tons (say around 35,000 pounds) if possible. Tracks are spaced 13 feet apart where practicable.

Motive Power

As of January 1, 1951, the Netherlands Railways owned:

744 steam locomotives
15 trolley electric locomotives
51 diesel electric locomotives
4 diesel mechanical locomotives
148 diesel electric "locomotors" (a locomotor is a small switching locomotive operated by one man, not necessarily a trained engineman)
10 heater cars
109 individual trolley electric motor cars
567 trolley electric bodies for 2-, 3-, 4-, and 5-car trains
157 diesel electric bodies for 3- and 5-car trains

This inventory has been augmented in 1951 and 1952 by the addition of Alsthom 2,950-horsepower 4-axle and of 3,800-horsepower 6-axle trolley electric locomotives, also some 6-axle Heemaf locomotives.

When these orders are completed the railroad will have 95 trolley electric locomotives. The number of streamlined trolley electric trains is also being increased gradually and some additional diesel electric motor trains are now being received.

Figure 5. Heemaf Co-Co trolley electric locomotive

Steam Locomotives

The steam locomotives are heterogeneous in type, ranging from relatively small switchers to Consolidations (2-8-0) and 10-wheelers (4-6-0). There are also a few heavier locomotives (used for such purposes as humping) and some 4-cylinder locomotives. Present plans call for the abandonment of all steam locomotive operations in the Netherlands as rapidly as possible.

Trolley Electric Locomotives

With the completion of the trolley electric locomotives now on order, the Netherlands Railways locomotive inventory will be shown in Table III.

The Oerlikon, Alsthom, and Heemaf locomotives are shown in Figures 3, 4, and 5. These locomotives are used for both passenger and freight work and do not leave the country because the German railways to the east are not electrified and the Belgian railroads to the south use 3,000 volts d-c.

Trolley Electric Motor Cars and Trains

Most of the passenger traffic is served by multiple-unit streamlined trains, groups of cars being permanently coupled to form train sets, Figure 6, and provided with automatic tight lock couplers at each end of each train set so that they may be operated in multiple. Passengers may circulate freely between cars of any train set, but there is no passage through the rounded ends of coupled sets. Any 2-, 3-, 4-, or 5-body set may be coupled and operated in multiple with any other set. These train sets are provided with second class and third class accommodations, baggage compartments, toilets, and in some cases with a buffet. There are also 34 streamlined postal cars (rounded nose at each end) arranged for coupling between or behind trolley electric train sets as needed. These postal cars are not motored.

Before the war the Netherlands Railways had 275 bodies for these trains. By 1951 approximately 220 of these had been recovered and rehabilitated. Others are being built, and by the end of 1952 it is expected that the trolley electric streamlined train inventory will be 241 streamlined train sets and 752 car bodies, as follows:

99	2-body train sets	198 car bodies
33	3-body train sets	99 car bodies
90	4-body train sets	360 car bodies
19	5-body train sets	95 car bodies

In addition, there are 101 old individual multiple-unit motor cars and 131 trailers not of streamlined construction which are still in use, these having been carried along from the initial electrification of 1927. This inventory will not be increased because these cars have almost outworn their usefulness and the streamlined train sets are so much more practical.

The trolley electric motor cars have been purchased at intervals and there are some marked variations in construction. The two basic types, of course, are the nonstreamlined individual units (1927) and the more modern streamlined trains of permanently coupled units adopted as standard in 1934. Although all of the latter are of the same general design, consisting of two, three, four, or five permanently coupled car bodies having an ellipsoid nose at each end of each train set, the trucks, body constructions, and equipment have varied from time to time with the acquisition of additional train sets.

The control systems of all of these cars embody bridging-type main circuit connections and automatic acceleration. Starting tractive forces are limited to give no more than 10-per-cent adhesive factor between the wheels and rails so that it is not necessary to carry sand on the cars. Two-car and 3-car train sets have one control equipment and four motors, while 4-car and 5-car sets use two equipments and eight motors, see Figure 7.

Diesel-Engined Motor Cars

The Netherlands Railways have recovered and rehabilitated but three of their single, mechanical drive, diesel cars; 17 of their 3-body diesel electric streamlined train sets, Figure 8; and 12 of their 5-body sets. Two additional 5-body sets have been acquired since the war. They also have on order 40 new 2-body train sets and 20 single car multiple-unit diesel electric cars. The latter will have the engines mounted under the car floors, whereas the engines of the earlier streamlined train sets had the engines mounted in a walled-off compartment of one of the

Table III. Netherlands Railways Inventory

Design	Oerlikon	Alsthom	Heemaf	Alsthom
Number of locomotives	10	50	25	10
Number series	1000	1100	1200	1300
Wheel grouping	1A—Bo—A1	Bo+Bo	Co—Co	Co—Co
Total weight, tons	100	80	120	110
Adhesive weight, tons	72	80	120	110
Maximum miles per hour	100	75	87	75
Horsepower (continuous)*	3,800	2,540	3,000	3,800
Miles per hour at continuous horsepower*	67	50	45	50
Wheel diameter, inches	61	49.3	43.4	49.3
Rigid wheelbase, feet and inches	7—3$^1/_4$	11—8	15—6	15—11

* At 1,350 line voltage

Figure 6. Four-car multiple-unit trolley electric train set

car bodies. It is expected that the diesel-engined passenger motor car inventory at the end of 1952 will be:

- 3 single cars, mechanical drive
- 20 single cars, electric drive, underfloor engines
- 40 2-body diesel electric train sets, underfloor engines
- 17 3-body diesel electric train sets
- 14 5-body diesel electric train sets

Four types of engines are used for the motor car trains. All engines (except the underfloor engines) for one train are placed in one body, usually with a baggage compartment at the ends of this car. The 3-body units are powered by two 400-horsepower (gross) engines and the 5-body trains by three 650-horsepower (gross) plants. All of the trains are arranged for multiple operation with any other diesel electric trains. The trains on order will have 6 horsepower per ton of train weight.

Diesel-Engined Locomotives

The Netherlands Railways use diesel and diesel electric locomotives for switching purposes and they also have a large number of diesel electric locomotors scattered over the system for handling local switching work. These locomotors are of 23 tons weight (United States 2,000 pound-tons) on two driven axles, each driven by an electric motor. The locomotor is provided with a cab, but there is also a running board on each side from which the movements of the unit also may be controlled. These locomotors are spotted at strategic locations where local switching must be done. Any local employee may operate the locomotor and trained enginemen seldom are used. This employee starts the diesel engine, switches and spots cars at local industries, gathers empties and loads as necessary for pickup by the next freight train, then puts the locomotor away and proceeds with his other duties. In normal weather, the operator usually rides on one of the running boards where he can get off easily to couple or uncouple cars. By minimizing delays to the freight trains and doing the local and industrial switching with one or (at the most) two unskilled employees,

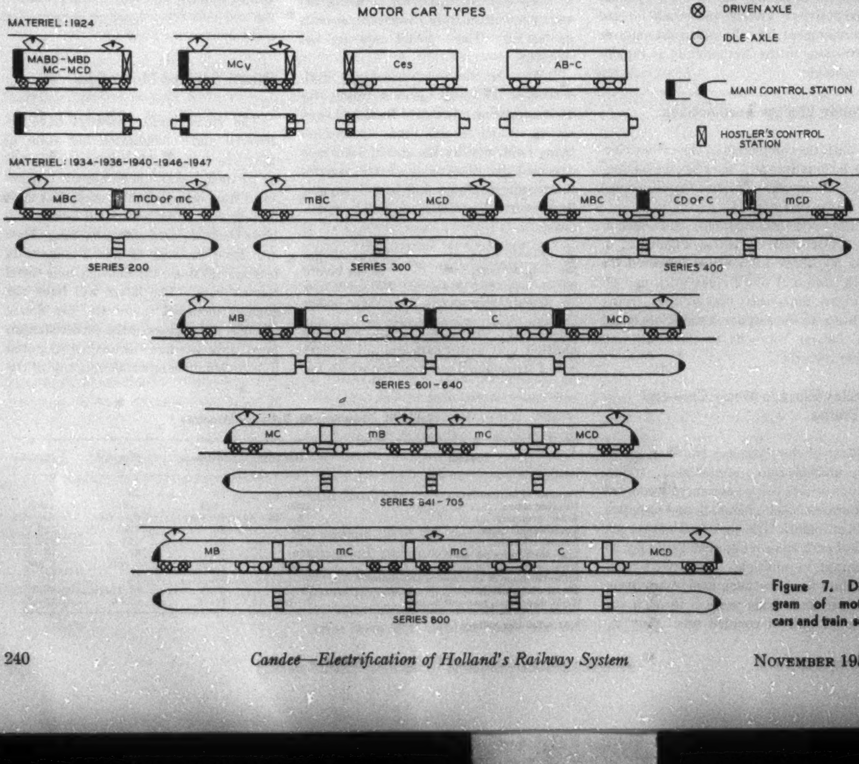

MOTOR CAR TYPES
MATERIEL: 1924
MABD-MBD
MC-MCD
MCv
Ces
AB-C
DRIVEN AXLE
IDLE AXLE
MAIN CONTROL STATION
HOSTLER'S CONTROL STATION
MATERIEL: 1934-1936-1940-1946-1947
MBC
mCD of mC
SERIES 200
mBC
MCD
SERIES 300
MBC
CD of C
mCD
SERIES 400
MB
C
C
MCD
SERIES 601-640
MC
mB
mC
MCD
SERIES 641-705
MB
mC
mC
C
MCD
SERIES 800

Figure 7. Diagram of motor cars and train sets

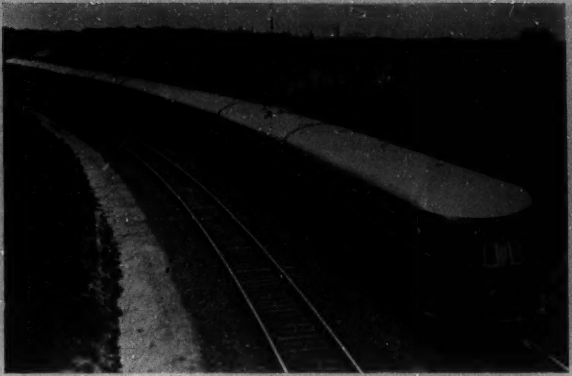


Figure 8. Three 3-body diesel electric train sets

the railroad has accomplished tremendous savings.

The inventory of diesel and diesel electric switching units now owned or on order is:

4 diesel locomotives, mechanical drive
50 English electric diesel electric switchers, type C
18 Whitcomb "war surplus" diesel electric switchers, type Bo-Bo
15 miscellaneous diesel electric switchers
163 locomotors, type Bo

Miscellaneous Rolling Stock

To complete the rolling stock picture, the railroad owns nearly 25,000 freight cars of various types, most of these being 2-axle construction with roller bearings and clasp brakes predominating on the later rolling stock. The inventory consists of box cars for fruits and vegetables, refrigerator cars, gondolas, container cars, and many other types as found on any railroad for normal or special purposes. The conventional European couplers are used with side buffers, the face of the right-hand buffer being convex and that of the left-hand being flat.

The railroad owns nearly 700 passenger cars for locomotive hauled trains. There are also 53 postal cars, 701 baggage cars (including high-speed freight cars), and 13 combination postal-baggage cars for such trains.

There are no signal lines through the trains. Signal for the departure of a train, then, is given by the station master with his red cap who blows a whistle and holds up a baton surmounted by a round disc. The driver, who has stepped out of the driving compartment or leans far out of the window, then goes back to his cab and starts the train. All trains (except steam locomotives) use but one man at the head end. For electric locomotive hauled trains it is customary to raise two pantograph trolleys to reduce the heating at the trolley wire under the heavy electric current drafts necessary for starting. With motor car trains, a sufficient number of pantographs are always kept up to care for the starting currents.

Trolley Electric Power Supply

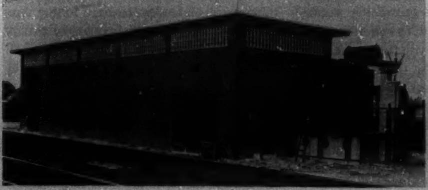
Power for the operation of trolley electric locomotives and multiple-unit trains is generated by the provincial and municipal coal burning power companies within the Netherlands. Three-phase 50-cycle power is delivered to the substations of the railroad at 10,000 volts, although some of the earlier substations are fed at 25,000 volts.

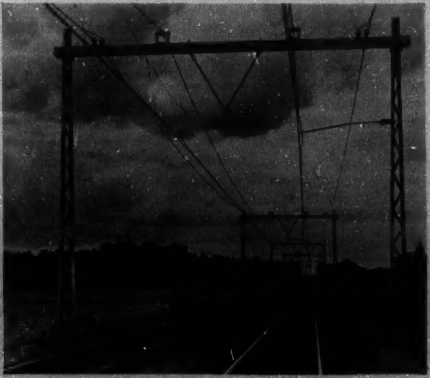
Power lines are generally underground and enter the substations in that way; see Figure 9.

Power is converted from 3-phase 50-cycle alternating current to direct current by mercury-arc rectifiers, the sizes in use being 500 kw, 1,000 kw, and 1,200 kw. Stations are built with one or two such units (three or four of the 500-kw size) as the local traffic requires, the older rectifiers generally being of the water-cooled, pumped type, some with 12 anodes and some with 6, while for the later extensions some are of the single-tank type with air cooling, some consist of a group of water cooled ignitrons, and others are sealed steel tanks with 6 anodes. In all, approximately 119,000 kw in rectifier capacity is now or will be installed by the end of 1952. The normal distance between substations is approximately $12^{1}/_{2}$ miles, and there will be a total of 71 substations at that time in addition to which there are three portable substations. Between substations are switching cabins for sectionalizing the lines, both the substations and the switching cabins being remotely controlled.

Figure 9 (below). Netherlands Railways substation

Figure 10 (right). Double track overhead contact system





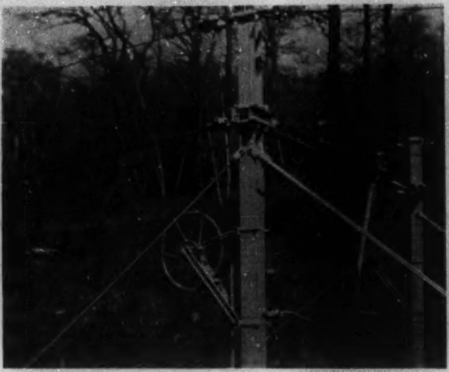


Figure 11. Contact wire tension device

Supporting Structures for the Contact Wires

The contact system of the Netherlands Railways has reached a high degree of practicability and relatively low cost. Starting with single poles and bracket arms on each side of double tracks, the railroad soon learned that poles would cant in the soft ground under wind and ice loads. The solution was to supply a cross member between the pole tops and eliminate the bracket arms. Thereafter there were installations of lattice-type poles, then of vertical I beams, then an ogive concrete arch, then vertical cylindrical uprights linked by a horizontal beam which type has become the simple standard system, Figure 10. Supporting structures were first spaced 75 meters (246 feet) apart, later changed to a spacing of 65 meters (213 feet), then finally standardized at 70 meters (230 feet).

Contact System

The standard contact system comprises a single copper catenary wire of 150 square millimeters' cross section (around 300,000 circular mils) with hangers from this to two horizontal and parallel contact wires, each with a cross sectional area of 100 square millimeters (each is approximately equivalent to a 4/0 wire). Paralleling these, mounted on insulators above the supporting cross member, is a single feeder wire with a cross section of 150 millimeters. The combined area of the four conductors per track, then, is 500 square millimeters (.0775 square inch or 987,000 circular mils). The contact systems of parallel tracks are tied together to improve the current carrying capacity. The average height of the contact wires above the surface of the rails is about 5.5 meters (18 feet 1 inch) with a maximum of 5.85 meters (19 feet 2 inches) and a minimum of 4.85 meters (15 feet 10 inches).

Modern practice for double track lines is to suspend a Y structure from the middle of the cross span of the supporting structures with a pulloff from this to the trolley wires, alternating at adjacent supports so that first the right-hand and then the left-hand trolley wires are pulled 35 centimeters off center (about $13^3/_4$ inches), first to one side and then to the other. When one pair of trolley wires is pulled toward the center by the center Y, the other is pulled in the same direction toward the side by a pulloff fastened to the vertical supporting pole. Thus, the trolley wires are staggered a total of 70 centimeters ($27^1/_2$ inches). These pulloffs also serve as braces against the strong winds prevalent in Holland. Trolley "pans" are fitted with lead-tin filled graphite bars across a width of 94 centimeters (37 inches) so that the contact is always made by these graphite bars.

The trolley wires are sectionalized approximately every 1,500 meters (4,921 feet). With these trolley wires anchored in the middle of the section, a tension device (Figure 11) is provided at each end, exerting a pull of 2,000 kilograms (4,410 pounds) or 1,000 kilograms (2,205 pounds) per trolley wire. This tensioning device consists of a pair of pulleys mounted on one of the supporting structures, one pulley being four times the diameter of the other, with a 500-kilogram weight suspended by a cable wrapped over the large pulley. The cable attached to the trolley wires is wound on and anchored to the smaller pulley. This forms a simple and practical tensioning arrangement.

Signals

At the present time, hand-operated signals are widely used throughout the densely settled, heavy traffic areas, although automatic color light signalling has been applied to some of the main lines. In general, there are but two signals—a distant and a home signal, spaced approximately 1,000 meters (3,280 feet) apart. Semaphore arms generally are used with the hand-operated signals, the outer ends of these arms indicating the type of signal: approach home, turnout, and so forth. Since fogs are frequent and heavy at some times of the year, the Netherlands Railways uses fog boards along the side of the track at a fixed distance ahead of each signal location. These slant upward at an angle and are striped for visibility. Stop signals cannot be passed even after a train comes to a full stop. The driver pulls up to the signal, blows his whistle, and then waits until the signal clears. Flagmen do not protect the rear of the train, so the driver starts immediately when the signal clears, which avoids considerable delay.

Shops and Maintenance Points

The Netherlands Railways has four main shops for the overhaul of equipment: Haarlem, in the northwestern area, handles trolley electric and diesel electric motor car overhauls; Tilburg, in the south central part of the country, takes care of locomotives (steam, trolley electric, and diesel electric); Amersfoort, just below the Zuider Zee in the central area, handles freight car repairs; and passenger cars are overhauled at Utrecht, in the central area. The continuing acquisition of trolley electric and diesel electric motor cars and locomotives and the shrinking of the steam locomotive inventory have forced changes in shop layouts and equipment and will require still further changes and expansions in the future.

Regular inspections and running repairs of the trolley electric rolling stock are now made at three points: Amsterdam (in the northwest) Leidschendam (central west), and Maastricht (in the extreme southeast). In general, Amsterdam takes care of trolley electric motor cars and locomotives, Leidschendam of trolley electric motor cars, and Maastricht of all types. This, of course, does not cover steam power, whose maintenance points are scattered over the system and are being reduced rapidly in number. Diesel electric motor cars are inspected at Utrecht, while diesel electric locomotives are inspected at Rotterdam (Fijenoord), Utrecht, Zwolle, Amsterdam, and Eindhoven.

The Netherlands Railways has a very effective inspection and overhaul program. As an instance of this, trolley electric motor cars operate in pools, say, 25 4-car trains in pool G, 13 3-car trains in pool H, and so forth. Each of these pools takes care of a set of regular runs, each train of the pool operating on a different run each day until all runs have been covered, then starting over again. A marker is placed in the front window to show the pool and the run, such as H-1. In this schedule of rotation, periodic visits to the inspection sheds are considered as one of the regular assignments. Thus, each of the trains performs exactly the same mileage as the others and is inspected at the same intervals. These trains are withdrawn once a year for light overhaul, once every 3 years for medium overhaul, and once every 10 years for complete rebuilding. Locomotives and diesel electric motor cars have similar assignments and periodic overhauls.

Conclusion

The decision of the Netherlands Railways to retain and to extend the 1,500-volt d-c system of electrification for their lines and services indicates an appreciation of the various factors which have an influence on the economies of this form of power. Thus, they have balanced the low maintenance expenses of medium voltage d-c motor car and locomotive equipment against a higher installation cost of the distribution network with close substation spacing (as compared to higher voltage systems), and find a net economy in their favor. In addition, the line troubles are very much lower with 1,500-volt direct current impressed upon the overhead network than if higher voltages (3,000 volts d-c or elevated alternating voltage) were employed. They also have reduced their substation and line costs to their practical essentials. As a result, the Netherlands Railways stands as one of the best examples of a 1,500-volt d-c electrification to be found anywhere in the world.

Discussion

D. R. MacLeod (General Electric Company, Erie, Pa.): A. H. Candee's paper describes the successful operation of the 1,500-volt d-c electrification of the Netherlands Railways. The United States also has an example of efficient and economical performance of a 1500-volt d-c system on the Illinois Central. This all-suburban-car electrification has been in operation since 1926. Each motor car has a 1-hour rating of 1,000-horsepower capable of hauling a semipermanently coupled trailer at speeds up to 70 miles per hour. The tendency on more recent installations on other suburban electrifications has been to go to all motor cars to give greater operating flexibility and to reduce the per cent adhesion during acceleration below 10 per cent. In the case of the Illinois Central, the decision to use a motor car and trail car combination was probably influenced by the existence of 45 suburban cars which could be used with slight modifications as trailers at the time that the lines were electrified. Combinations of motor cars with one or more trailers are generally considered less flexible than single motor cars especially where traffic varies widely during the day. A failure on a motor car takes the entire group of cars out of service; a road failure on a train of cars consisting of a single motor car and trailers does not permit the train to proceed under its own power. Presumably these factors are not as important in Holland as in this country. The large number of different train sets consisting of different numbers of cars permanently coupled together is unusual.

As manufacturers of electrical equipment, the General Electric Company would like to see more railroad electrification. The Netherlands have achieved standardization on one system within her boundries. This is a worth-while accomplishment. Each system has its inherent advantage but none of the advantages are as great as the advantage of mass production. Railroad engineering, mechanical and electrical, is a matter of constant trial and error so that the more there are of the same kind to try, the more progress will be made in eliminating the errors. Time and again it has been shown that lower costs are achieved by manufacturing large quantities of the same equipment that gives broad coverage to the requirements of a large number of users rather than by developing a unique solution for each one. Nevertheless, until a mass market develops under the domination of one of the many systems now in use, we have to plan for all of them. It is well to recognize, therefore, that until standardization changes the picture, the cheapest locomotive and car equipments are those that use 600 volts direct current collected from an overhead trolley or third rail; that 1,500 volts direct current runs 600 volts a close second, and 3,000 volts direct current is not far behind. Whether they are for commercial frequencies or for 25 cycles a-c locomotives and cars cost more per pound of weight on drivers than any of the d-c locomotives and cars that we have built to date. Since the Netherlands Railways only has approximately 3,000 miles of track to electrify and has such a large motor-car traffic, a 1,500-volt d-c electrification should give them a low cost and efficient railway system. They have made a worth-while contribution to the cause of electrification.

T. F. Perkinson (General Electric Company, Schenectady, N. Y.): In view of the current world-wide interest in commercial-frequency railroad electrification, generated to a large degree by the French National Railways' experimental installation at Annecy, it is of interest to note that the Netherlands Railways decided to rebuild their war-damaged lines with a relatively low-voltage d-c contact system. Undoubtedly those responsible for the decision considered commercial frequency as a possible means of meeting requirements, and it is apparent that, for the special and particular conditions surrounding the Netherlands system, the 1,500-volt d-c system appeared to be the most acceptable solution.

The author's observation that the predominance of electric motor-car trains in the Netherlands system carried considerable weight in the decision to retain the 1,500-volt d-c system, suggests that, where vehicle density is relatively great, as is the case in heavy suburban transport operations, relatively low-cost low-voltage d-c propulsion equipment has considerable merit. The economies realized might well more than offset the added costs involved for wayside equipment with closer substation spacing and heavier distribution systems.

Because of the limited geographical area and relatively short distances involved, the necessity of considering the future extension of electrification over lines of light traffic and long distances apparently was not influencing in the choice of system.

While comparisons at times are odious or otherwise undesirable, one cannot but help to compare the use of 1,500 volts in the Netherlands with the adoption of 3,000 volts as a national standard in the neighboring country of Belgium. Both countries are similar in geographical location, area, and population, and the uninitiated, at least, would assume that rail transportation requirements in the two countries are somewhat similar. True, the original electrification at 1,500 volts in the Netherlands antedated the adoption of 3,000 volts as a standard in Belgium, and the urge to retain the system with which they were most familiar, coupled with the opportunity for salvage of some of the prewar equipment, may have been the impelling reasons for the retention of the 1,500-volt system in the Netherlands. Apparently there was no consideration of any moment with respect to possible interchange of motive power between the two countries, when the decision to re-establish the 1,500-volt system in the Netherlands was made.

The relatively low height of the contact wire for a standard-gauge system is worthy of comment. While the height of the contact wire may have but a remote influence on the choice of system, it is of interest to note that the maximum height of 19 feet 2 inches in use on the Netherlands Railways would have a sensible influence on the initial cost of the distribution-system supporting structures. This height contrasts with dimensions of 22 feet for the Illinois Central 1,500-volt and 24 feet for the Lackawanna 3,000-volt suburban electrifications in the United States. Undoubtedly the latter heights are necessitated by the requirements of clearance to high freight-loading gauges encountered in American railroad-ing—extremities that apparently are not practiced on the Netherlands lines.

The decision of the Netherlands to retain 1,500 volts direct current as a national standard furthers the belief that currently there is no one system of electrification available that is best suited to all conditions, as some proponents of particular systems would have us believe. A proper regard for many influencing factors—traffic density, the nature of the service to be performed existing installations with their concomitant capital investments, the desirability of interchange of motive power with neighboring systems, future traffic prospects, primary power supply availability, and established national standards are some of the factors which must be taken into account in prescribing electrification for any particular piece or system of railroad if the most economical and practicable solution to the problem is to be attained.

A. H. Candee: Since preparing this paper, comments have been received from Mr. J. P. Koster, Chief Engineer of the Traction and Rolling Stock Department of the Netherlands Railways, which are essentially as follows:

"You give a short outline of the reasons which induced the Netherlands Railways to continue the use of 1,500 volts direct current. Actually this might also be stated in a negative way, in that we could find no advantage in the use of 3,000 volts direct current and the use of high voltage alternating current was entirely out of the question. In those days (end of 1945) we found that by adopting 3,000 volts direct current rather than 1,500 volts direct current, the saving in depreciation of capital costs would be nil, although the saving in energy consumption would be approximately 150,000 guilders per year at a coal price of 15 guilders per ton. This saving, however, would have been heavily overbalanced by the increase in rolling stock maintenance. Also, with 3,000 volts on the trolley line in our climate, it would not have been possible to do the trolley-line maintenance work without deenergizing the wire, which would be of considerable inconvenience from an operational standpoint.

"In reference to point 4 under 'Re-selection of the 1,500-Volt D-C System', I would like to emphasize that the number of electric vehicles per kilometer of electrified line is rather high, so that it pays to choose an electrification system wherein the first cost of propulsion equipments is low. The higher price of fixed installations for a lower voltage d-c system may be accepted as it will be compensated for by the lower first costs of the propulsion equipments. The relationship of investment for the Netherlands Railways (electrified lines) is approximately: rolling stock 35 per cent; fixed power installations 30 per cent; and permanent way and signals 35 per cent.

"You mention under 'Diesel-Engined Motor Cars' the equipment which we will have at the end of 1952. We now find that the 20 single cars and the forty 2-body sets will not get in service before June 1954. The seventeen 3-body train sets have now been increased to 29.

"In connection with multiple operation of diesel-electric trains, we can operate six power plants from one driver's cab, which means two 5-body or three 3-body train sets or combinations of these.

"Under 'Trolley Electric Power Supply' you give a total rectifier capacity of 119,000 kw. By May of 1953 we expect to have 132,000 kw installed in a total of 74 substations.

"In your description of the 'Contact System,' you mention that trolley pans are fitted with lead-tin filled graphite bars. This applies only to motor cars as we use copper-carbon strips for locomotives. These have a better heat conductivity than the lead-tin filled strips although they wear faster.

"In regard to shops, the Utrecht workshop has now been abandoned, with the steel cars now being overhauled at Haarlem and the older wooden stock (soon to be scrapped) overhauled at Amersfoort. We have now reduced the number of steam locomotive inspection points to nine."

Analogue Methods for Optimum Servomechanism Design

F. C. FICKEISEN
NONMEMBER AIEE

T. M. STOUT
ASSOCIATE MEMBER AIEE

Synopsis: Optimum design of servomechanisms based on the criterion of minimizing the integral of squared error for representative inputs has been discussed by Hall,[1] Phillips,[2] and others. Investigations of related criteria, the mean-square and mean-absolute error for a square-wave input, are described in this paper. Experimental results on simulated second- and third-order servomechanisms check closely with analytical values, and the optimum adjustments obtained in this way coincide with those calculated for step function inputs. Measurement and minimization of mean-absolute error appears particularly promising as an experimental design technique, since standard rectifier-type voltmeters can be used and the resulting optimum responses conform with subjective standards of good response.

A COMPLETE description of the transfer characteristic of a linear servomechanism is given by a statement of the response to one input—impulse, step, or sinusoid. With this information, the response to any other input can be calculated.

Significant aspects of the response can be given in less detailed form in a variety of ways, by rise time and overshoot for step function inputs, peak magnification and bandwidth for sinusoidal inputs, steady-state error for ramp inputs, or by error coefficients which are useful for more general inputs.[3] Several of these measures of response are generally considered in the process of writing specifications, picking components, and testing the finished servomechanism; although these quantities are interdependent, their interrelationships are not always easy to determine and such abbreviated descriptions of one response do not adequately define the others.

Various measures of system response have been discussed[1,2,4–11] which take account of the complete response to a specific input, usually the step function. These measures are integrals of the error, such as

$$I_1=\int_0^\infty e\,dt \tag{1}$$

$$I_2=\int_0^\infty |e|\,dt \tag{2}$$

$$I_3=\int_0^\infty t\,e\,dt \tag{3}$$

$$I_4=\int_0^\infty e^2\,dt \tag{4}$$

$$I_5=\int_0^\infty t\,e^2\,dt \tag{5}$$

$$I_6=\int_0^\infty t^2e^2\,dt \tag{6}$$

$$I_7=\int_0^\infty e^4\,dt \tag{7}$$

$$I_8=\int_0^\infty e^6\,dt \tag{8}$$

Of these, the first four are of particular interest. The integrated error, or "control area," is easy to calculate but has some serious shortcomings.[5,6] The integral of the absolute value of the error is more satisfactory as a measure of response and also more difficult to calculate. The weighted integral of the error has some desirable properties and is evaluated readily

---

Paper 52-246, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Pacific General Meeting, Phoenix, Ariz., August 19–22, 1952. Manuscript submitted March 24, 1952; made available for printing July 10, 1952.

F. C. FICKEISEN is with the Boeing Airplane Company, Seattle, Wash. T. M. STOUT is with the University of Washington, Seattle, Wash.

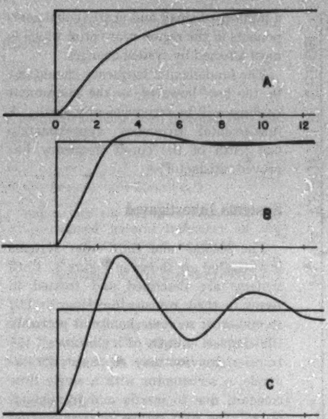


Figure 1 (left). Typical servomechanism response to step-function input

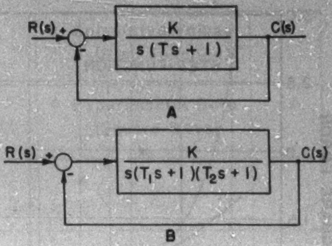


Figure 3 (right). Block diagrams of servomechanisms investigated

A—Second-order system
B—Third-order system

by operational methods, using the system transfer functions.[6] The integral of squared error has been used extensively in the optimum design of circuits[11] being convenient analytically and adaptable to noise problems; Phillips[2] gives a complete discussion of its peculiarities as a measure of system response.

Measures of this type are particularly useful in determining the optimum adjustment of a given system or in choosing between systems with similar characteristics; the "best" system is the one that minimizes the integral employed.

Proposed Procedure

Because of the difficulties involved in determining optimum adjustments with these criteria by analytic methods, it was decided to find them experimentally. With present-day emphasis on the use of analogue computers in servomechanism design, it appeared that an investigation of the applicability of such a procedure for simple systems might provide a basis for interesting studies of more complicated systems.

The integrals of absolute and squared error for a square-wave input were selected for investigation, for several reasons:

1. A square wave approximates the step function, for which mean-square optimizations of second- and third-order systems already have been determined.
2. These integrals are measured easily by averaging thermocouple- and rectifier-type voltmeters.
3. These measures of response are expected to lead to system adjustments in accord with subjective standards of good response.

The investigations described in this paper are an extension of work described by Carleton,[12] who employed the criterion of minimum integrated squared error in determining the optimum parameters of a 2-stage Rototrol exciter voltage regulating system by analogue methods. A more elaborate study of optimum filters, using similar measuring techniques, has been carried out by Stutt.[13]

Characteristics of Method

Typical servomechanism responses to step function inputs are shown in Figure 1, and a square-wave response corresponding to Figure 1(C) is considered in greater detail in Figure 2. The error, defined as

$$e(t)=r(t)-c(t)$$

is shown in Figure 2(B). The shaded area in Figure 2(B), averaged over a cycle, is the reading of a half-wave rectifier-type instrument. The squared error, whose average is indicated by a thermocouple instrument, is shown in Figure 2(C).

General characteristics of the minimum mean-absolute and mean-square error criteria can be deduced from Figure 2. Since large errors are weighted heavily by

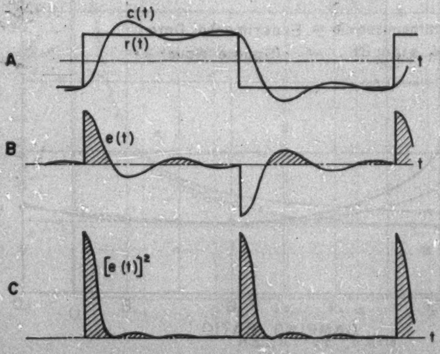


Figure 2. Typical servomechanism response to square-wave input

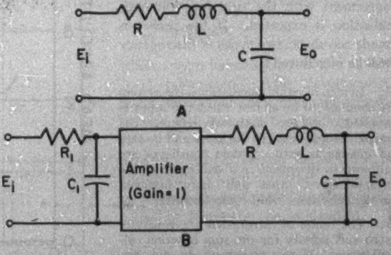


Figure 4. Equivalent networks for servomechanisms of Figure 3

A—Second-order system
B—Third-order system

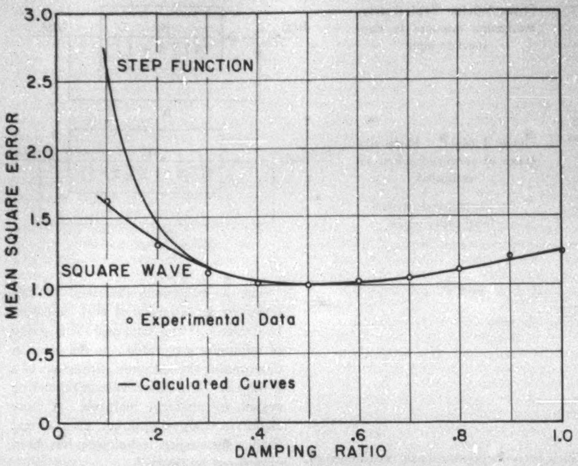


Figure 5. Mean-square error as a function of damping ratio for a second-order servomechanism

the mean square error criterion, emphasis is placed on rapid reduction of the large and inevitable error which follows each abrupt change in the input, even at the expense of considerable overshoot and continuing oscillation. The mean absolute error criterion, on the other hand, weights all errors equally and will lead to systems with a less rapid initial error reduction, less overshoot, and fewer oscillations.

Selection of Square-Wave Frequency

A square-wave test may be regarded as a repeated examination of the step function response of the system. Viewed in this light, the square-wave period chosen should be long enough for transient oscillations to disappear in a half cycle, if good agreement with the step function optimization is expected. If the period is made too short, two types of discrepancy will be introduced:

1. In step function calculations, the system is assumed to be at rest when the input is applied; because of inertia effects, the initial error is equal to the input and its rate of change is zero. Unless transients are allowed to disappear in a half-cycle of the square wave, other and more involved initial conditions will be encountered.

2. The limits of integration are generally zero and infinity for the step function calculations. If the transients do not disappear in a half-cycle, some of the integral is lost.

These errors are clearly most important in systems with either too little or too much damping.

This type of test also may be considered a simultaneous probing of the frequency response characteristics at a number of different frequencies. Since the magnitudes of the frequency components of a square wave vary inversely with frequency, it appears desirable to have the fundamental frequency of the square wave small compared to the cutoff frequency or bandwidth of the system, thus placing a number of small and nearly equal components in the region near cutoff which is most affected by system changes.

The fundamental frequency should not be too low, however, or the instrument readings will be inconveniently small. A fundamental frequency approximately one tenth of the cutoff frequency has proved satisfactory.

Systems Investigated

The second- and third-order systems investigated are shown in Figure 3. Both systems are described and treated in terms of their transfer functions.[3] The second-order servomechanism is probably the simplest system of importance; the transfer function may represent, for example, a servomotor with a single time constant due to inertia and friction effects. The next degree of complication is indicated in the third-order system by another time constant which might be traced to the amplifier controlling the motor. Although shown as single-loop systems, they may be interpreted as simplifications of multiple-loop systems of the type encountered with tachometer feedback, or as approximations to some higher-order systems.

For the second-order system, the closed-loop function may be expressed as

$$\frac{C(s)}{R(s)}=\frac{\omega_0^2}{s^2+2\zeta\omega_0 s+\omega_0^2} \tag{9}$$

where

$\zeta=(4KT)^{-1/2}$, the damping ratio
$\omega_0=(K/T)^{1/2}$, the undamped natural frequency

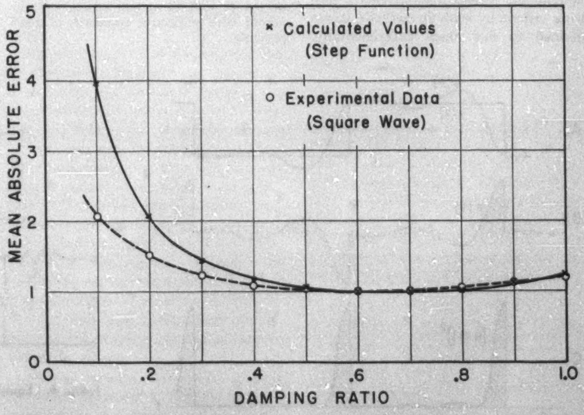


Figure 6. Mean-absolute error as a function of damping ratio for a second-order servomechanism

The transfer function of the third-order system

$$\frac{C(s)}{R(s)}=\frac{K}{T_1T_2s^3+(T_1+T_2)s^2+s+K} \tag{10}$$

is assumed to have been factored into the form

$$\frac{C(s)}{R(s)}=\frac{\omega_c}{(s+\omega_c)}\times\frac{\omega_0^2}{(s^2+2\zeta\omega_0 s+\omega_0^2)} \tag{11}$$

ζ and ω_0 are the damping ratio and undamped natural frequency of the second-order factor, as before

ω_c is the "break frequency"[14] or $-3DB$ frequency of the first-order factor

Unfortunately, there are no convenient formulas for ζ, ω_0, and ω_c in terms of T_1, T_2, and K; the solution is not difficult, however, in numerical cases.

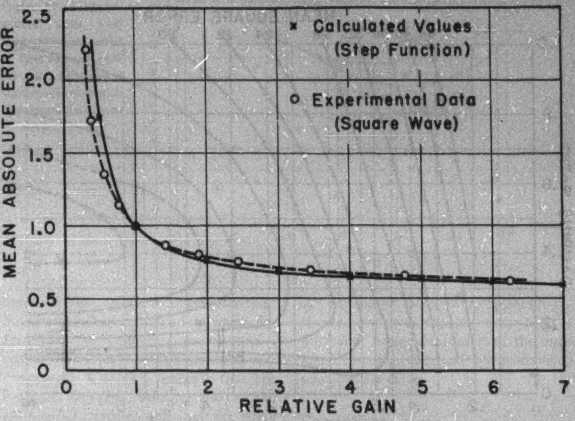


Figure 8. Mean-absolute error as a function of gain for a second-order servomechanism

Bandwidth

Servomechanisms, viewed in the frequency domain, are low-pass filters which are conveniently described in terms of gain and phase characteristics. Bandwidth is an important parameter, since speed of response varies inversely with bandwidth. A specification of the required rise time for a step function input thus indicates the minimum allowable bandwidth. However, because of noise, it is generally desirable to employ the least bandwidth commensurate with the prescribed transient behavior.[15] For this reason, bandwidth was regarded as an invariant quantity and emphasis was placed on those system adjustments which do not affect bandwidth.

Ignoring bandwidth reduces by one the number of parameters required to characterize a system. In the second-order system, attention is thus focused on the effects produced by varying the damping ratio ζ. For the third-order system, the parameters chosen for study are ζ, the damping ratio of the second-order factor, and ω_c/ω_0, a dimensionless frequency ratio which locates the first-order factor along the frequency scale relative to the undamped natural frequency of the other factor. Since this ratio is close to unity for the cases of interest, ω_0 again becomes a measure of bandwidth.

No generality is lost by this procedure, as is apparent from the following argument. The error for an nth-order system can be expressed in terms of a dimensionless time parameter $\omega_0 t$ in which t is real time and ω_0 is a measure of system bandwidth, and $(n-1)$ dimensionless parameters a_i which are either damping ratios or frequency ratios. For a step function input, the integral of any function of the error, $F(e)=E(\omega_0 t, a_1, a_2, \ldots, a_{n-1})$, is then

$$I=\int_0^\infty E(\omega_0 t, a_1, a_2, \ldots, a_{n-1})\,dt \tag{12}$$

$$=\frac{1}{\omega_0}\int_0^\infty E(\omega_0 t, a_1, a_2, \ldots, a_{n-1})\,d(\omega_0 t) \tag{13}$$

$$=\frac{1}{\omega_0}I(a_1, a_2, \ldots, a_{n-1}) \tag{14}$$

where I is, as indicated, a function of only $(n$-$1)$ parameters. This development, which shows that these integrals vary inversely with bandwidth regardless of the effects of the other parameters, is merely another way of expressing the fact that, in the absence of noise and other parameters remaining constant, better response is obtained by increasing the bandwidth.

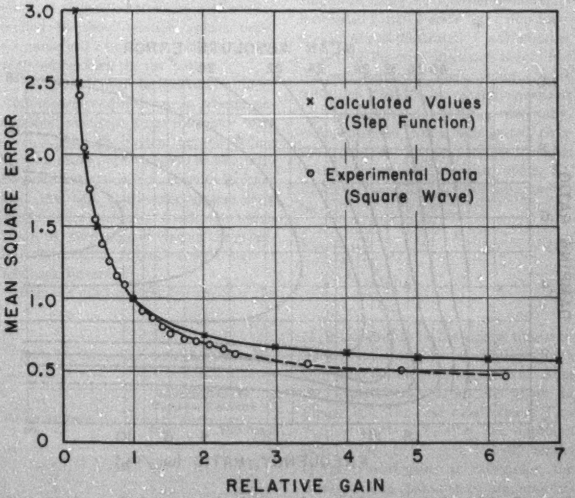


Figure 7. Mean-square error as a function of gain for a second-order servomechanism

Experimental Technique

For experimental studies, the second- and third-order systems were represented

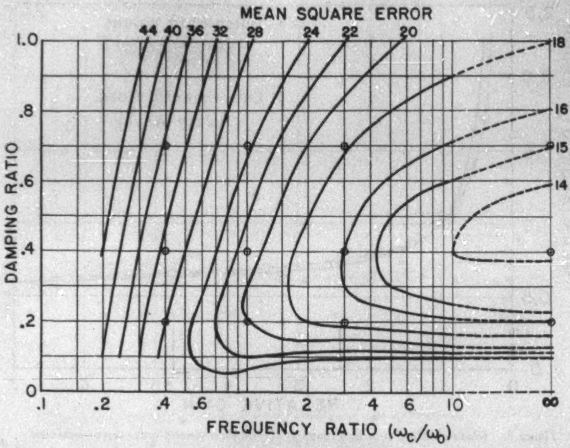


Figure 9. Mean-square error contours for a third-order servomechanism as functions of damping ratio ζ and frequency ratio

by the equivalent electric networks of Figure 4. (Although discussed in servomechanism terms, the responses of these networks may be interesting in their own right.) Inputs were supplied by a standard square-wave generator. Voltmeters were connected to measure the difference between E_i and E_o which is analogous to the servomechanism error. The rectifier-type voltmeter was a standard 1,000-ohm-per-volt a-c instrument; the mean-square voltmeter was a thermocouple instrument driven by a difference amplifier, isolated by cathode-followers to prevent loading of the network. The effects of variations in the various servomechanisms parameters were examined by changing the values of the corresponding network elements.

The parameters of the second-order servomechanism, expressed in terms of the circuit elements, are

$$\omega_o = \frac{1}{\sqrt{LC}} \tag{15}$$

$$\zeta = \frac{R}{2}\sqrt{\frac{C}{L}} \tag{16}$$

In keeping with equation 14, ω_o was fixed and ζ was varied by changing R, which includes any resistance in the inductor and generator. Values of the circuit elements were checked by noting that, for a sinusoidal input whose frequency is ω_o, the phase shift is 90 degrees independent of damping and the output-input ratio is $1/(2\zeta)$. Effects of gain variations were studied by changing C, since C is analogous to $1/K$.

The network representation of the third-order system follows naturally from the factored transfer function of equation 11. The characteristic frequency of the added resistance-capacitance section, $\omega_c = 1/(R_1C_1)$, was varied by changing R_1. The isolating amplifier was a cathode-follower whose loss was accounted for by comparing E_o with a fraction of E_i.

Treating ω_o as a constant, curves of mean-square and mean-absolute error were obtained using ζ as the independent variable and ω_c/ω_o as a parameter. Effects produced by varying the gain were obtained indirectly, by translating changes of gain into corresponding changes of ζ, ω_c, and ω_o, and using the data already obtained.

Results

The results of the analytical and experimental investigations are summarized in Figures 5–11.

For a second-order system, the integrated squared error for a step function input is[1]

$$I = \frac{1}{\omega_0}\,\frac{1+4\zeta^2}{4\zeta^2} \tag{17}$$

Values obtained from this equation are plotted in Figure 5 with a curve of calculated mean-square error for a square-wave input and corresponding experimental data. The square-wave calculations and measurements shown are based on a fundamental frequency ω_f equal to one fourth of ω_o, the undamped natural frequency, and exhibit the expected deviation from the step function curve for low damping ratios. Better agreement is observed with $\omega_f = \omega_o/10$.

Similar curves of mean absolute error as a function of damping ratio are given in Figure 6.

Using the mean-square error criterion,

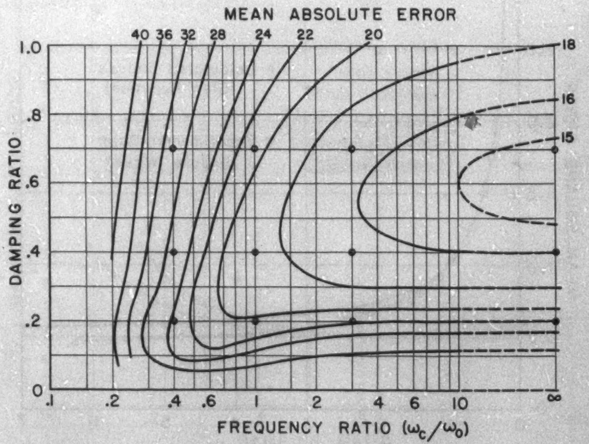


Figure 10. Mean-absolute error contours for a third-order servomechanism as functions of damping ratio ζ and frequency ratio

the optimum damping ratio is found to be approximately $\zeta=0.5$, in agreement with the analytical value. For mean-absolute error, the measured value is approximately $\zeta=0.6$, in contrast with an expected value of about 0.7; as the minimum of the curve is extremely broad, this difference is not considered important. The response obtained on a mean-absolute error basis is less oscillatory than the response obtained by minimizing mean-square error and agrees somewhat better with intuitive or subjective standards of good response; the result is, for example, more nearly in accord with the common practice of adjusting galvanometers, oscillographs, accelerometers, and other similar devices to have a damping ratio in the range of 0.6 to 0.7.

Results obtained by varying the gain of the second-order system are shown in Figures 7 and 8. The calculated curves were derived from the analytical data already on hand by noting that ζ varies inversely, and ω_o directly, with $K^{1/2}$. A mean error of one unit was assumed for $\zeta=0.7$ and $K=1$, and successive corrections applied for changes in ζ and ω_o as K was varied. As expected,[1] there is no true optimum value of gain, since the increase in bandwidth more than compensates for the reduction in damping ratio.

Data for the third-order system was regarded as defining surfaces of mean-square and mean-absolute error as functions of the two variables ζ and ω_c/ω_o; equal error contours for these surfaces are presented in Figures 9 and 10. The units used are somewhat arbitrary and are intended to show relative values only; no numerical comparisons should be attempted between the two plots.

Two general observations can be made. Both measures of error are, in general, decreasing functions of the frequency ratio ω_c/ω_o. Since the system becomes effectively a second-order system as this ratio becomes large, this fact seems to indicate that the best third-order system is a second-order system, a conclusion which is not too surprising in view of the generally detrimental effects of time lags in feedback systems.

It is not always possible to reduce the order of the system transfer function, however, so there is some interest in determining the optimum parameters of a system in which two time constants are important. The plots indicate that the optimum damping ratio increases as the ratio ω_c/ω_o increases and that, for small damping ratios, an optimum ω_c/ω_o can be found. As before, mean absolute error optimizing tends to favor systems with more damping.

Table I. Comparison with Hall's Results

Frequency Ratio (ω_c/ω_o)	Damping Ratio for Optimum Response	
	Hall	This Paper
0.6	0.17	0.15
1.0	0.32	0.25
2.0	0.41	0.33
∞	0.5	0.5

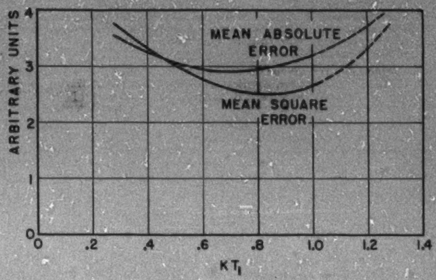


Figure 11. Mean-square error and mean-absolute error as functions of gain for a third-order servomechanism with two equal time constants ($T_2=T_1$)

Hall[1] investigated the third-order system using a different set of parameters but did not make an exhaustive search for the optimum combination. His results, based on the integrated squared error for a step function input, have been converted into the parameters used in this paper and compare as shown in Table I.

Aigrain and Williams[16] have given a method for the design of optimum video amplifiers, using the criterion of minimum integrated squared error for a step function input and illustrated by a third-order system of the type discussed here. For the case considered, they obtain $\omega_c/\omega_o=0.566$ and $\zeta=0.138$, in general agreement with the results obtained experimentally.

An optimum gain is anticipated, since the third-order system becomes unstable if the gain exceeds a critical value, K_c, which is found by Routh's criterion[2] to be

$$K_c=\frac{1}{T_1}+\frac{1}{T_2} \tag{18}$$

$$=\frac{1}{T_1}\left[1+\frac{T_1}{T_2}\right] \tag{19}$$

(T_1 is assumed to be the largest time constant.) Calculated curves of mean-square and mean-absolute error, derived from the experimental data, are given in Figure 11. The case considered is $T_2=T_1$, for which $K_cT_1=2$. The optimum gain is found to be seven to eight tenths of the critical gain; in this range, the corresponding parameters are approximately $\omega_o=0.65/T_1$, $\zeta=0.25$, and $\omega_c=2.5\omega_o$. As noted previously, optimum response on a mean-absolute error basis is somewhat better damped than the optimum response obtained by minimizing mean-square error.

Conclusions

Analytical investigations and experimental studies using simulated systems have indicated that square-wave tests employing mean-square and mean-absolute error measurements are a satisfactory means for determining the optimum parameters of simple servomechanisms, provided that a true optimum exists. The mean-absolute error, because it is more easily measured and leads to systems with slightly more damping, seems to be particularly promising.

Although the tests were conducted on network analogues of the servomechanisms, it appears feasible to apply the same general procedure to actual servomechanisms. Low-frequency square-wave generators are now commercially available and instruments with the necessary characteristics probably can be obtained.[17]

References

1. The Analysis and Synthesis of Linear Servomechanisms (book), A. C. Hall. Technology Press, Cambridge, Mass., 1947.

2. Theory of Servomechanisms (book), H. M. James, N. B. Nichols, R. S. Phillips. McGraw-Hill Book Company, New York, N. Y., 1947.

3. Servomechanisms and Regulating System Design (book), H. Chestnut, R. W. Mayer. John Wiley and Sons, Inc., New York, N. Y., 1951.

4. The Dynamics of Automatic Controls (book), R. C. Oldenbourg, H. Sartorius. American Society of Mechanical Engineers, New York, N. Y., 1948.

5. A Note on Control Area, T. M. Stout. *Journal of Applied Physics*, American Institute of Physics (New York, N. Y.), volume 21, November 1950, pages 1129–31.

6. Some Design Criteria for Automatic Controls, Paul T. Nims. *AIEE Transactions*, volume 70, part 1, 1951, pages 606–11.

7. The Calculation of the Optimum Parameters for a Following System, C. Mack. *Philosophical Magazine* (London, England), volume 40, September 1949, pages 922–28.

8. An Improved Pneumatic Control System, **R. E. Claridge.** *Transactions*, American Society of Mechanical Engineers (New York, N. Y.), volume 73, April 1951, pages 297–305.

9. Controller Settings for Optimum Control, **W. A. Wolfe.** *Transactions*, American Society of Mechanical Engineers (New York, N. Y.), volume 73, May 1951, pages 413–18.

10. Theoretical Considerations on the Optimum Adjustment of Regulators, **P. Hazebroek, B. L. van der Waerden.** *Transactions*, American Society of Mechanical Engineers (New York, N. Y.), volume 72, April 1950, pages 309–15.

11. Extrapolation, Interpolation, and Smoothing of Stationary Time Series (book), **N. Wiener.** John Wiley and Sons, Inc., New York, N. Y., 1949.

12. The Transient Behavior of the Two-Stage Rototrol Main Exciter Voltage Regulating System as Determined by Electrical Analogy, **James T. Carleton.** *AIEE Transactions*, volume 68, part I, 1949, pages 59–63.

13. Experimental Study of Optimum Filters, **C. A. Stutt.** *Technical Report Number 182*, Research Laboratory of Electronics, Massachusetts Institute of Technology (Boston, Mass.), May 15, 1951.

14. Principles of Servomechanisms (book), **G. S. Brown, D. P. Campbell.** John Wiley and Sons, Inc., New York, N. Y., 1948.

15. The Analysis and an Optimum Synthesis of Linear Servomechanisms, **Donald Herr, Irving Gerst.** *AIEE Transactions*, volume 66, part II, 1947, pages 959–70.

16. Design of Optimum Transient Response Amplifiers, **P. R. Aigrain, E. M. Williams.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 37, August 1949, pages 873–79.

17. Indicating Instruments at the Servomechanism Frequencies, **Wilson S. Pritchett, Robert M. Saunders.** *AIEE Transactions*, volume 69, part II, 1950, pages 980–85.

Discussion

Paul T. Nims (Chrysler Corporation, Detroit, Mich.): Mr. Fickeisen and Professor Stout are to be congratulated on achieving a successful marriage of the analytical derivation of optimum response with the electric systems of the analogue computers which are so often used instead of analytical methods in connection with servomechanism design. While it has long been the practice to plot out the response of a servomechanism repeatedly as one or more of its parameters are varied and then to examine these curves for "best response," this procedure has a tendency to fill filing cabinets with response curves. Using the new procedure proposed by the authors, it will only be necessary to record one or two meter readings in connection with each set of parameters investigated, and both the time required for testing and the record keeping problem should be significantly reduced.

In some physical systems, which might be tested by the authors procedure, it is inconvenient to excite the system with repeated square waves so that a method for getting the same numerical information from a single step function input might be of interest. In the testing of permanent magnets, it is customary to use a ballistic galvanometer to give a measure of the total flux of a magnet from the voltage induced in a search coil which has passed over the magnet. This is possible because the ballistic galvanometer gives an almost steady deflection which is proportional to the integral of the current passing through it. The control area (equation 1 of the paper) thus might be evaluated simply by applying the error signal to high resistance galvanometer and exciting the system with a single-step function. Obviously the integral of the absolute value of the area (equation 2 of the paper) might be obtained by interposing a suitable rectifier between the error signal and the ballistic galvanometer. As the authors point out, the control area and similar unweighted measures of the error, such as the rms error, all tend to overemphasize the unavoidably large error at time zero so that a very desirable criterion of the response might be the weighted control area which is defined in references 2 and 6 of the paper. The instrumentation for the quantity t times the error is a little more complicated but might be well worth the trouble. For analogue computers which operate at somewhat near real time, like the Reeves and the Boeing computers, there are available servomultiplier units which could be fed with the error and with a signal increasing with time and which would give the product of these two as the output which could be connected to a ballistic galvanometer for reading the weighted control area. If a rectifier were used, then we would have the integral of the absolute value of t times the error and this has promise of being an especially useful measure of response of the servo as it should minimize the first overshoot somewhat more than the weighted error itself does.

As the authors point out, the methods of this paper can be applied to actual servomechanisms using either low-frequency square-wave generators or, as previously suggested, a single-step input. This not only gives a simple laboratory test of the servomechanism but might also be useful as an acceptance or routine inspection test.

Analogue Computers for Feedback Control Systems

ROBERT A. BRUNS
ASSOCIATE MEMBER AIEE

Synopsis: Feedback control design is passing into a third stage of development as the systems become more complicated, less subject to linear treatment, and more susceptible to noise problems. This article is a review of analogue computers which are useful in studying involved linear systems and those which do not lend themselves to direct mathematical analyses.

FEEDBACK control design is entering a third stage of development. Before 1934, empirical methods were satisfactory, since regulators required relatively little accuracy. In 1934, Hazen[1] published his work on servomechanisms, contemporary with Bush's article on the differential analyzer.[2] Operational methods of handling differential equations were used on simple systems. Laplace and Fourier transforms were applied, locus diagrams in the complex plane were used as a design tool, and pole-zero techniques proved useful. But the present emphasis is toward those systems which do not lend themselves readily to simple techniques of linear analysis and synthesis. This has been brought about by requirements of higher quality of response, simplicity, and physical limitations of components, such as saturation and backlash. Furthermore, the input data is not so clear-cut. Both external and internal noise may be present.

Thus, the present tendency in feedback control systems is to include the effects of saturation resulting from large-signal input and the irregularities of amplifying elements, and to utilize relay systems and nonlinear elements to obtain a better solution to the control problem. One typical example is to take the simple proportional positioning system and use a relay or variable-μ tube to change the damping coefficient as a function of the ratio of the input to the error signal. Such problems lead to a very complicated theoretical framework and the designer must turn back to semiempirical methods to supplement the mathematical tools he

---

Paper 52-247, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Pacific General Meeting, Phoenix, Ariz., August 19–22, 1952. Manuscript submitted January 21, 1952; made available for printing July 14, 1952.

Robert A. Bruns is with the University of California, Berkeley, Calif. The background for a portion of this article was obtained while the author was at the Hughes Aircraft Company Research and Development Laboratories, Culver City, Calif.

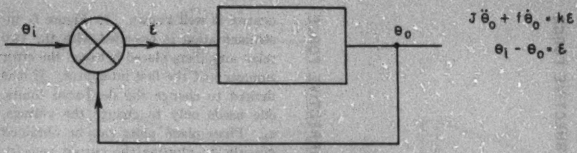

θ_i
ε
θ_0
$J\ddot{\theta}_0 + f\dot{\theta}_0 = k\varepsilon$
$\theta_i - \theta_0 = \varepsilon$

Figure 1. Linear feedback control system and equations

has available and to develop their usefulness.

One such prop is the electronic analogue computer, which has been developed to a point where its accuracy and reliability permit investigations of nonlinear systems. A typical computer is the REAC (Reeves Electronic Analogue Computer)[3] which allows solutions of sixth order differential equations with 0.2-percent error, compared with numerical solutions. Having a frequency response from 1/50 to 5 cycles per second, it is adequate for most work in feedback control systems.

Computers for Simple Linear Systems

To illustrate the method of instrumentation, consider the simple system of Figure 1. Solving for the value of $d^2\theta_0/dt^2$ (the second derivative of the output position or its acceleration) from the open-loop equation

$$\frac{d^2\theta_0}{dt^2} = -\frac{f}{J}\,\frac{d\theta_0}{dt} + \frac{k}{J}\epsilon \tag{1}$$

Since the integral of the acceleration is the velocity, the quantities on the right-hand side of equation 1 constitute the input to the first integrator, the output then being the velocity, $-d^2\theta_0/dt^2$; see Figure 2. If there were an initial velocity, this could be taken care of by an initial condition voltage applied to this integrator. The second integrator has θ_0 as its output. The closed-loop equation is accounted for in the summer. As systems become more complicated, opportunities for ingenuity present themselves. Needless to say, if the system is so complicated that instrumentation becomes a design problem, reexamination of the fundamental properties of the system under consideration may be in order.

Computers for Relay and Saturating Systems

The system of Figure 1 possesses an amplifier with the transfer characteristics of a linear system as shown in Figure 3(A). If saturation is present, a point of limiting constant corrective action is reached as in Figure 3(B). The characteristics of the amplifier of a relay system are shown in Figure 3(C).

To illustrate how an analogue computer may be used for a system like that in Figure 3(B), consider the case in which there is a limiting velocity in the linear system of Figure 1. The linear equations apply when the error signal is within the limits, $\pm\epsilon_0$. However, beyond these limits, the velocity is a constant, $\pm K$. Instrumentation is shown in Figure 4. A pair of relay amplifiers[4] are placed between the two integrators with the output contacts connected as shown. The contacts are in the plus position when the sum of the inputs to the relay amplifier is positive. Otherwise, they are in the opposite position. Since these relay amplifiers are high-gain amplifiers, definite action is achieved, much more so than with diode limiters. Also, fast relays are now available, having a response time of 100 microseconds, so there need be no sacrifice in speed of response. When $-\epsilon_0 \leqq -d\theta_0/dt \leqq +\epsilon_0$ the contacts are in positions shown and operation is the same as in Figure 6. However, when $-d\theta_0/dt > +\epsilon_0$, the sum of the inputs to *R.A.* 2 becomes positive, the direct line is broken as the *R.A.* 2 contacts switch to the plus sides and the voltage, $+K$ is applied to the input of the second integrator. On the other hand, when $-d\theta_0/dt < -\epsilon_0$, the center line is broken in the other place and $-K$ switched in. Initial conditions are indicated to complete the picture.

A simple relay servomechanism with friction and inertia in the output is shown in Figure 5, together with equations of operation. The amplifier action is like that of Figure 3(C). Recent improvements in relay design and the need for a physically small feedback control system help to justify their use, even in positioning systems. Their application in process

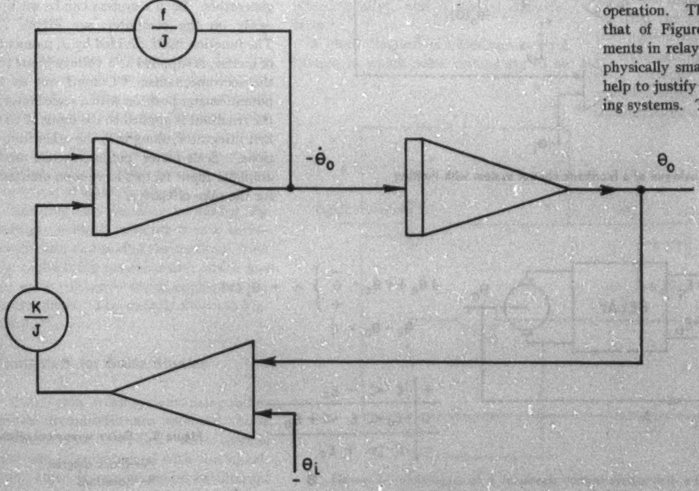


Figure 2. Analogue of linear feedback control system

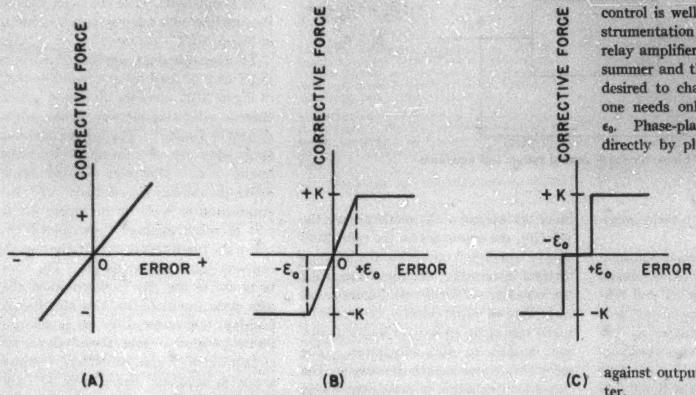


Figure 3. Error corrective action for

A—Linear system
B—Linear system with saturation
C—Relay system

control is well known. In Figure 6, instrumentation is provided with the two relay amplifiers placed between the error summer and the first integrator. If it is desired to change the dead-zone limits, one needs only to change the voltage, ϵ_0. Phase-plane plots can be obtained directly by plotting the output velocity against output position on an X-Y plotter.

Computers for Systems with Time-Varying Parameters

In certain feedback control systems, particularly those in which a radio link is present in the error-measuring element or between the command station and the error signal, one might be confronted with the open-loop equation

$$J\frac{d^2\theta_0}{dt^2}+f\frac{d\theta_0}{dt}=k(t)\epsilon \qquad (2)$$

where $k(t)$ is some function of time. If the function is mathematically expressible, it can be generated using analogue components. Otherwise, it can be supplied to the servomechanism from function generators. Such a system can be set up easily on the computer, see Figure 7. The function itself, divided by J, moment of inertia, is supplied as a voltage input to the servomechanism. Coming out as a potentiometer position with a scale factor, the resultant is applied to the input of the first integrator, along with the other functions. Scale-factor potentiometers and amplifier input factors have been omitted for the sake of clarity.

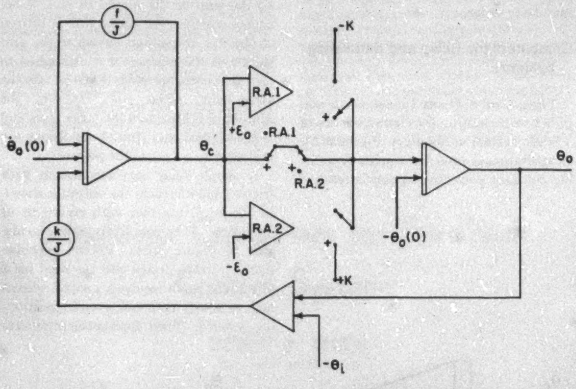


Figure 4. Analogue of a feedback control system with limiting

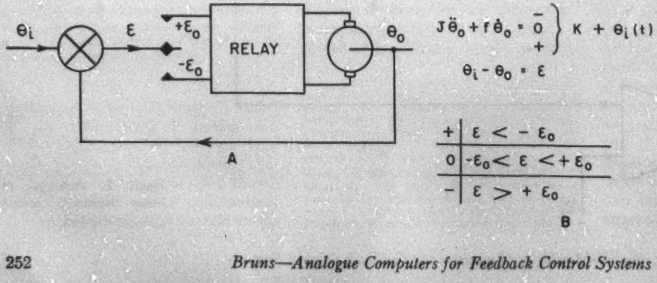


Figure 5. Relay servomechanism

A—Block diagram
B—Equations

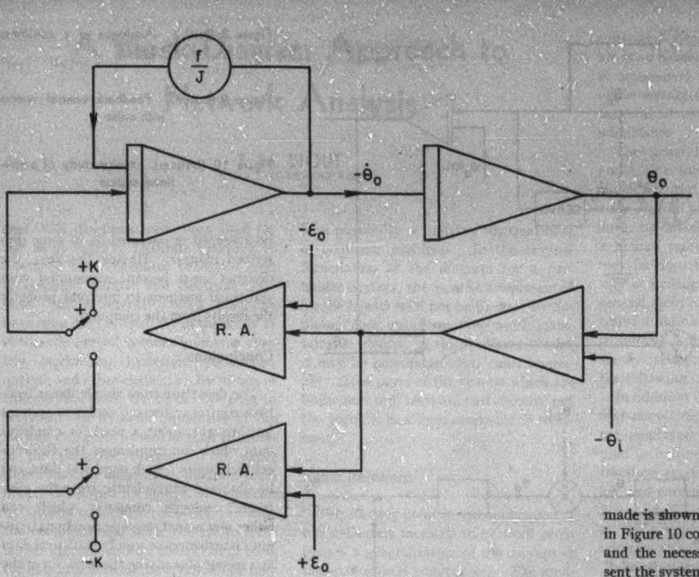


Figure 6. Analogue of relay servomechanism

Computers for Nonlinear Systems

If the output of a positioning system has a torque which varies as the square of the velocity, one must resort to special mathematical methods. Consider the system with the equations

$$J\frac{d^2\theta_0}{dt^2}+f\,\frac{d\theta_0}{dt}\pm B\,\frac{d\theta_0{}^2}{dt}=K\epsilon \tag{3}$$

$$\theta_i-\theta_0=\epsilon \tag{4}$$

The plus-minus sign indicates that the sense of torque depends upon the direction of velocity. When velocity is positive, the torque also has a positive sign in the equation. Reversing velocity changes the sign. This is necessary because the squared velocity term remains positive regardless of sense of velocity.

Squaring may be done by taking the voltage $-d\theta_0/dt$ applying it to a servomechanism and taking the resultant from the multiplying potentiometer which has the same voltage $-d\theta_0/dt$ applied across its terminals. The setup is shown in Figure 8.

Computers for Noise Studies

Once again, when systems using radio-carrier frequencies are required, noise problems must be studied. First, there is external noise combined with the signal; then, there is receiver noise or internal noise. The nature of the noise must be known before proper noise generators can be constructed. The building of noise generators is no simple problem and is still in its infancy. Sources of noise that are in use are thyratron tubes, crystal diodes, and photomultiplier tubes. One such generator uses a thyratron tube operated at low current density, a high-Q selective filter, and a special demodulator.[5]

A block diagram of a feedback control system in which noise studies are to be made is shown in Figure 9. Its analogue in Figure 10 contains two noise generators and the necessary components to represent the system. It is a nonlinear system, so that theoretical methods alone would scarcely allow meaningful analytical results. It is difficult enough to interpret results from an analogue computer. Usually, many runs of the problem are required. Root mean square and mean values of series of runs are of interest. To get numerical data, digital read-out devices are necessary. Automatic sequencers may be helpful in carrying out the many operations required without continuously resetting the problem on the computer. Information must be stored in order to get statistical data. Since it appears first as a voltage, digital devices must be available which change this voltage to a number stamped on tape which can then

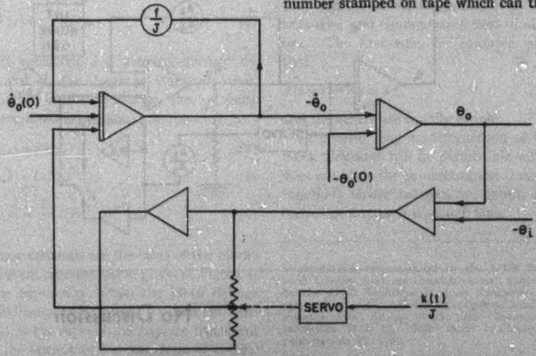


Figure 7. Analogue of a feedback control system with a time-varying parameter

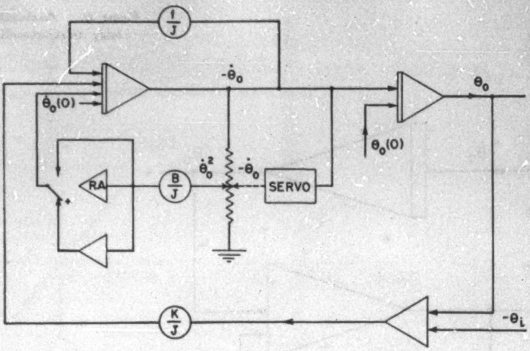


Figure 8 (left). Analogue of a nonlinear system

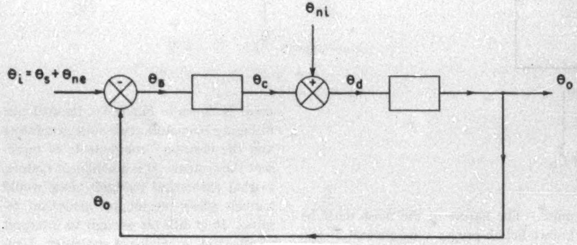


Figure 9 (center). Feedback control system with noise

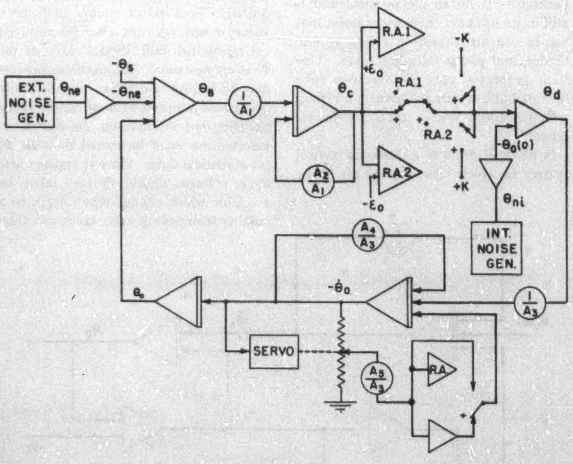


Figure 10 (bottom). Noise study of a nonlinear system

be averaged or operated on in some prescribed manner. Having the data, the designer must become acquainted with statistical methods to interpret properly the results from the computer.

Conclusions

The departure from simple linear feedback control systems to nonlinear systems has brought about a need for empirical data which can supplement the theoretical techniques. Such empirical data can be obtained conveniently from the electronic analogue computer which can either plot actual response functions to input disturbances or supply statistical data as a means of studying the behavior of the system in the presence of noise. The problem resolves itself not into whether it can be set up on the computer but whether it can be expressed mathematically. Thus the designer can turn to more basic problems,[6] such as a study of the nature of noise in a system, with the confidence that, regardless of how complicated the feedback control system under consideration, it can be instrumented on an analogue computer.

References

1. THEORY OF SERVOMECHANISMS, **H. L. Hazen.** *Journal,* Franklin Institute (Philadelphia, Pa.), volume 218, September 1934, pages 279–330.

2. A NEW TYPE OF DIFFERENTIAL ANALYZER, **V. Bush, S. H. Caldwell.** *Journal,* Franklin Institute (Philadelphia, Pa.), volume 240, October 1945, pages 255–326.

3. COMPACT ANALOGUE COMPUTER, **Seymour Frost.** *Electronics* (New York, N. Y.), July 1948, pages 116–22.

4. THE USE OF HIGH-SPEED RELAYS IN ELECTRIC ANALOGUE COMPUTERS, **R. R. Bennett, A. S. Fulton.** *AIEE Transactions,* volume 70, part II, 1950, pages 1553–55.

5. THE GENERATION AND MEASUREMENT OF LOW FREQUENCY RANDOM NOISE, **R. R. Bennett, A. S. Fulton.** Hughes Aircraft Company (Culver City, Calif.), August 1951.

6. CYCLONE PROJECT SYMPOSIUM I ON REAC TECHNIQUES (book). Reeves Instrument Corporation, New York, N. Y., 1951.

No Discussion

A Block-Diagram Approach to Network Analysis

T. M. STOUT
ASSOCIATE MEMBER AIEE

BLOCK diagrams have been used for many years to facilitate planning of assembly lines, chemical processes, and complicated electric and electronic systems. More recently, block diagrams have been used extensively in the study of automatic control systems involving electric, mechanical, hydraulic, pneumatic, optical, and other elements. By means of transfer functions, the blocks are used to represent the dynamic properties of the various elements, as well as their function in the complete system. This paper shows that the same block-diagram techniques developed for control systems can be applied to electric networks, reducing the computational labor needed to solve network problems and contributing fresh insights into network behavior which may be helpful in synthesis problems.

Although the block-diagram method can be used to establish network theorems, its main application is in the solution of real network problems, where it supplements the usual series-parallel circuit reductions, Y-delta transformations, and the standard network theorems. It is particularly adapted to the determination of transfer quantities, such as voltage ratios or transier impedances and admittances. The block-diagram manipulations used in network reduction eliminate the need for formal solutions of simultaneous equations; such calculations are replaced by block transformations which keep the physical system in view and which permit checking of the solution at each step. These manipulations can be carried out graphically if the algebraic representation of the blocks becomes too cumbersome.

A further advantage of the block-diagram approach is the fact that, since the method is based on operational mathematics, transient and steady-state solutions are obtained with almost equal facility. The block diagram of a system incorporates all of the necessary information for determining the response to any input; the choice of a constant, pulse, or sinusoidal input can be postponed until the network reduction has been completed. Several inputs and outputs may be considered at the same time.

Finally, the block-diagram methods offer a possibility of unifying the treatment of nonlinear problems. In the original formulation of the diagram for a particular system, the relations represented by the blocks need not be linear. By applying block transformations which maintain the identity of the nonlinear blocks, it may be possible in some cases to simplify those parts of the system which are linear and well behaved and thereby put the problem in a form amenable to treatment.

Basic Relations

The starting point in the application of block-diagram methods to network problems is a generalization of the concept of impedance (and admittance). The fundamental differential equations for the basic circuit elements—resistance, inductance, and capacitance—are

$$e_R(t) = R\, i(t) \qquad (1)$$

$$e_L(t) = LD\, i(t) \qquad (2)$$

$$e_C(t) = \frac{1}{CD}\, i(t) \qquad (3)$$

where D is a differential operator d/dt. The corresponding relations for sinusoidal voltages and currents may be written as follows

$$E_R(j\omega) = RI(j\omega) \qquad (4)$$

$$E_L(j\omega) = j\omega LI(j\omega) \qquad (5)$$

$$E_C(j\omega) = \frac{1}{j\omega C}\, I(j\omega) \qquad (6)$$

If equations 1–3 are transformed by means of the Laplace transformation, ignoring initial conditions, the following operational expressions are obtained

$$E_R(s) = RI(s) \qquad (7)$$

$$E_L(s) = LsI(s) \qquad (8)$$

$$E_C(s) = \frac{1}{Cs}\, I(s) \qquad (9)$$

These relations are the basis of the block-diagram representation given in Figure 1. The expression within the block defines the "transfer function"[1] which, in general, is the ratio of the Laplace transform of the output to the transform of the input. The blocks imply a unidirectional, cause-and-effect relation: a current flowing in an element produces a voltage which is determined by use of an impedance, and a voltage applied to an element produces a current determined by use of an admittance.

The generalized impedances R, Ls, and $1/Cs$ evidently can be obtained from the differential equations by substituting s for D, and from the usual a-c impedances by substituting s for $j\omega$. Thus, in a limited sense, the Laplace variable s may be regarded merely as a shorthand way of writing D or $j\omega$. Taking a more general view, however, use of the Laplace transformation techniques makes possible a unified approach to transient and steady-state problems, as will be seen in the following.

In addition to the blocks used to represent circuit elements, two other symbols are employed. The "summing point," used to represent the additions and subtractions encountered in Kirchoff's voltage and current laws, is shown in Figure 2(A) and (B). The plus and minus signs at the summing point indicate whether addition or subtraction is intended. The "take-off point," shown in Figure 2(C), signifies that the same signal may proceed along several different paths to be applied to several different blocks. As suggested by the figure, both voltages and currents may be regarded as signals.

Block Transformations

The rules for the transformation of block diagrams have been summarized by Graybeal;[2] the most important of these rules are given in Figure 3.

Of the rules for transforming block diagrams, only the last two which deal with feedback loops need to be memorized; the others can be derived as needed, since they are based on the associative, distributive, and commutative laws of algebra. The first rule, for example, says that

$$A(x \pm y) = Ax \pm Ay \qquad (10)$$

and the others are equally simple.

In general, any transformation of the block diagrams will be permissible which does not alter the product of the transfer functions either between an input and output or around a feedback loop.

Paper 52-254, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Pacific General Meeting, Phoenix, Ariz., August 19–22, 1952. Manuscript submitted May 19, 1952; made available for printing July 21, 1952.

T. M. Stout is with the University of Washington, Seattle, Wash.

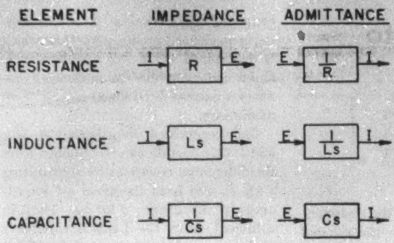
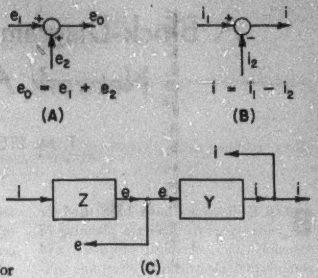

ELEMENT
IMPEDANCE
ADMITTANCE
RESISTANCE
I
R
E
E
1/R
I
INDUCTANCE
I
Ls
E
E
1/Ls
I
CAPACITANCE
I
1/Cs
E
E
Cs
I

Figure 1 (left). Block - diagram representation of basic electric-circuit elements using generalized impedance and admittance

Figure 2 (right). Summing points and take - off point


e₁
+
+
e₀
e₂
e₀ = e₁ + e₂
(A)
i₁
+
−
i
i₂
i = i₁ − i₂
(B)
i
i
Z
e
e
Y
i
i
e
(C)

Simple Circuits

As an introduction to the solution of network problems by block-diagram methods, consider the single-section ladder network consisting of two resistances shown in Figure 4(A). The equations for this circuit, assuming that e_0 is measured by a voltmeter which requires no current from the network, are

$$i=\frac{e-e_0}{R_1} \tag{11}$$

$$e_0=R_2 i \tag{12}$$

The unorthodox statement of this problem is chosen because it illustrates a general method for treating ladder networks. The block diagram Figure 4(B) may be reduced in two ways. Using the rule 8 for general feedback loops Figure 4(C) is obtained directly, from which the final result shown in Figure 4(E) may be obtained by rule 5. On the other hand, rule 3 can be used to derive the diagram of Figure 4(D) from the original diagram; rule 7 then produces the final diagram.

A network which is slightly more complicated is given in Figure 5(A). The combination of resistance R_2 and capacitance C may be treated in two ways, depending on the manner in which the current equation is written. If we write

$$i_2=i_1-i_C \tag{13}$$

the block diagram of Figure 5(B) results. Or, if we write

$$i_C=i_1-i_2 \tag{14}$$

Figure 5(C) is obtained. Application of rule 8 to the inner feedback loop in either diagram gives the diagram of Figure 5(D), from which Figure 5(E) is obtained by use of rules 5 and 7.

If the capacitance in the network of Figure 5(A) is allowed to approach zero, the network reduces to the resistance network of Figure 4(A); this change is accompanied by the disappearance of the inner feedback shown in Figure 5(B). Likewise, if R_2 is made infinite, the inner loop in Figure 5(C) disappears and an ordinary resistance-capacitance circuit results. It might be noted that the larger block in Figure 5(D) represents the

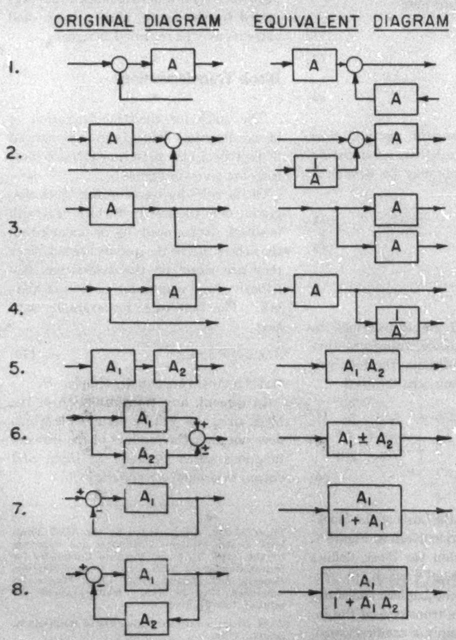
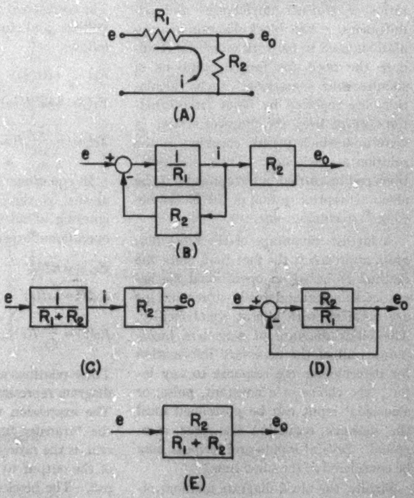

ORIGINAL DIAGRAM
EQUIVALENT DIAGRAM
1.
A
A
A
2.
A
A
1/A
3.
A
A
A
4.
A
A
1/A
5.
A₁
A₂
A₁ A₂
6.
A₁
A₂
+
±
A₁ ± A₂
7.
+
−
A₁
A₁/(1 + A₁)
8.
+
−
A₁
A₂
A₁/(1 + A₁ A₂)

Figure 3 (left). Rules for transforming block diagrams

Figure 4 (below). Resistance network and associated block diagrams


R₁
e
e₀
R₂
i
(A)
e
+
−
1/R₁
i
R₂
e₀
R₂
(B)
e
1/(R₁+R₂)
i
R₂
e₀
(C)
e
+
−
R₂/R₁
e₀
(D)
e
R₂/(R₁ + R₂)
e₀
(E)

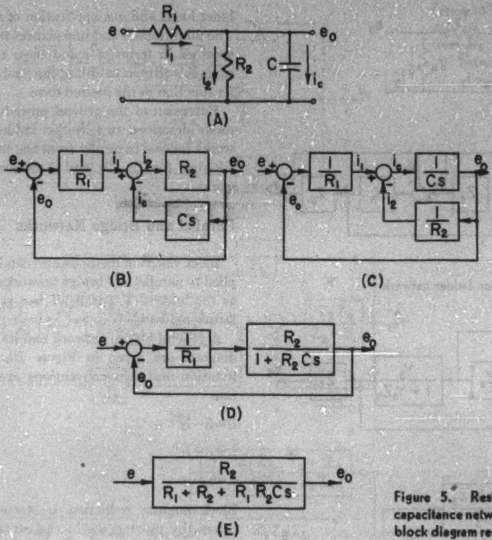

(A)
(B)
(C)
(D)
(E)
R_1
R_2
C
e
e_0
i_1
i_2
i_c
$\frac{1}{R_1}$
Cs
$\frac{1}{Cs}$
$\frac{1}{R_2}$
$\frac{R_2}{1+R_2Cs}$
$\frac{R_2}{R_1+R_2+R_1R_2Cs}$

Figure 5. Resistance-capacitance network and block diagram reduction

equivalent impedance, expressed in operational or general form, of R_2 and C in parallel.

Ladder Networks

The previous examples, which could have been treated more directly by conventional methods, were presented in order to show how block-diagram manipulations can be used to solve network problems. The usefulness of this procedure is shown more clearly by the examples to follow.

Consider first the 2-section, resistance-capacitance ladder network shown in Figure 6(A). The block diagram, which is drawn directly from the circuit diagram using the procedures suggested in the previous examples, is given in Figure 6(B). Inspection of the block diagram reveals that all of the necessary network equations are correctly represented.

Reduction of the diagram is started by moving the block containing R_1 to the left of the first summing point and then interchanging the first and second summing points, and moving the block containing C_2 to the right of the last take-off point; the resulting block diagram is Figure 7(A). The two inner loops are then reduced to single blocks, as shown in Figure 7(B), by use of rule 8. The remaining reductions are routine and the final result is given in Figure 7(D).

If the two sections are identical, then $R_1=R_2=R$ and $C_1=C_2=C$; the final transfer function is

$$\frac{e_2(s)}{e(s)}=\frac{1}{R^2C^2s^2+3RCs+1} \tag{15}$$

which is known to be the correct result.[3]

Consider next the 3-section ladder network of Figure 8(A). Contrary to the usual practice, the series elements are described by their admittances and the shunt elements by their impedances; the block diagram illustrated in Figure 8(B) is most readily constructed on this basis.

Two steps in the reduction of this network are shown in Figure 9(A) and (B). The general procedure, developed in the previous examples, is again employed: blocks are moved past summing and take-off points in such a way that the order of these points may be interchanged, forming inner loops which may be reduced by rules 7 and 8 to give new blocks on which the same process may be repeated. Further reduction of Figure 9(B) is seen to require application of rule 7 to each of the

Figure 6 (below). Block diagram and 2-section ladder network

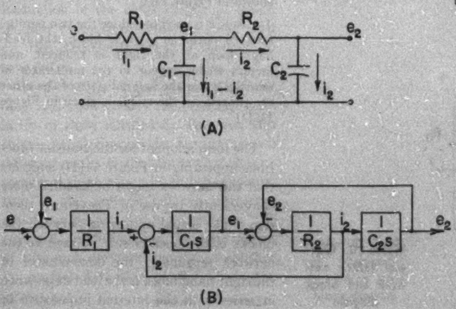
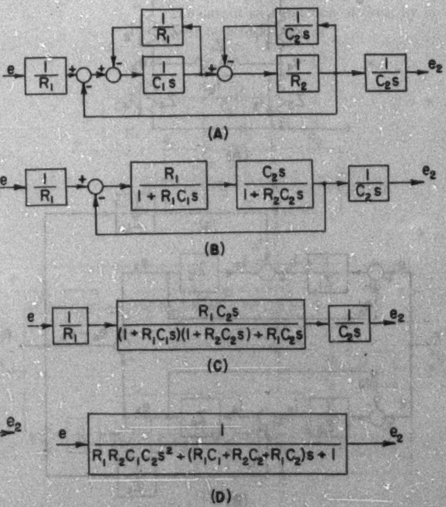

(A)
(B)
R_1
R_2
e_1
e_2
i_1
i_2
C_1
C_2
i_1-i_2
e
$\frac{1}{R_1}$
$\frac{1}{C_1s}$
$\frac{1}{R_2}$
$\frac{1}{C_2s}$

Figure 7 (right). Block-diagram reduction for 2-section ladder network


(A)
(B)
(C)
(D)
e
e_2
$\frac{1}{R_1}$
$\frac{1}{C_1s}$
$\frac{1}{R_2}$
$\frac{1}{C_2s}$
$\frac{R_1}{1+R_1C_1s}$
$\frac{C_2s}{1+R_2C_2s}$
$\frac{R_1C_2s}{(1+R_1C_1s)(1+R_2C_2s)+R_1C_2s}$
$\frac{1}{R_1R_2C_1C_2s^2+(R_1C_1+R_2C_2+R_1C_2)s+1}$

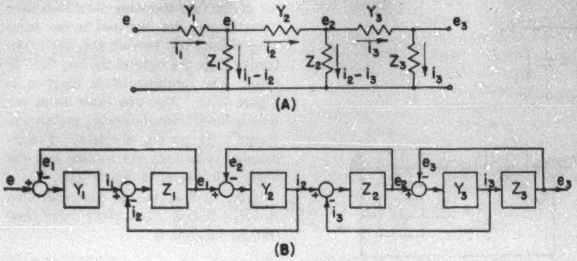


Figure 8. Block diagram and 3-section ladder network

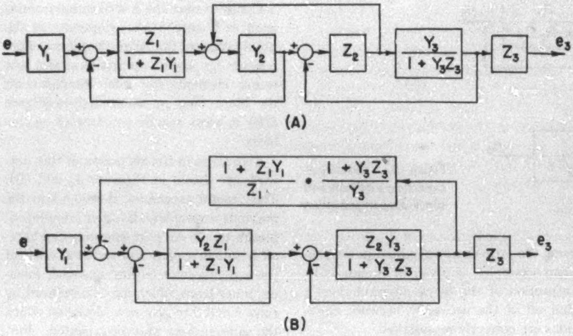


Figure 9. Partial reduction of the block diagram for 3-section ladder network

inner loops and one application of rule 8. The admittances and impedances may be specified in terms of s and these expressions substituted to obtain the final transfer function in the desired form.

Extension of the general procedure already described to n-section ladder networks follows naturally from the method adopted for the 2- and 3-section networks.

Parallel and Bridge Networks

Block diagram methods also can be applied to parallel and bridge networks, such as the bridged-T, parallel-T,[4] or general-lattice network.

A general-lattice network and its block diagram are given in Figure 10. The notation and sign conventions are such that

$$i_3 = i_1 - i_L \qquad (16)$$

$$i_4 = i_2 + i_L \qquad (17)$$

$$e_0 = e_a - e_b \qquad (18)$$

Block-diagram reduction is started, as before, by moving the blocks containing $1/Z_1$ and $1/Z_2$ to the left of the first summing points and then interchanging the order of the summing points in each of the forward paths. The inner feedback loops thus obtained are reduced to single blocks, giving Figure 11(A). The remaining steps are:

1. The block containing Z_3 and Z_4 is moved to the right, past the summing point; in accordance with rule 2, the reciprocal transfer function is inserted in the upper feedback loop.
2. The block diagram is redrawn in an equivalent form Figure 11(B) to suggest more clearly the next steps.
3. The largest block on the main forward path is moved to the left of the first summing point and into the upper feedback path, and the order of the summing points is reversed. The minor feedback loop including Z_3, Z_4, and Z_L is reduced to a single block. The result of these two steps is shown in Figure 11(C).
4. Rule 6 is used to reduce the two parallel forward paths to a single block, the feedback loop at the right is reduced, and factors which appear in the numerator of one block and the denominator of the other are cancelled; the result is given in Figure 11(D).

The form adopted for the transfer functions appearing in Figure 11(D) suggests that this network might be handled more conveniently by use of Thevenin's theorem. The signal appearing between the blocks is the open-circuit voltage at the network terminals; the denominator of the right-hand block is the load impedance in series with the internal impedance of

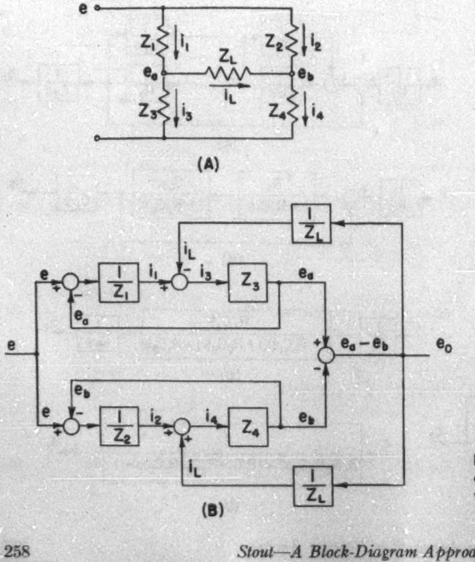


Figure 10. General lattice network and block diagram

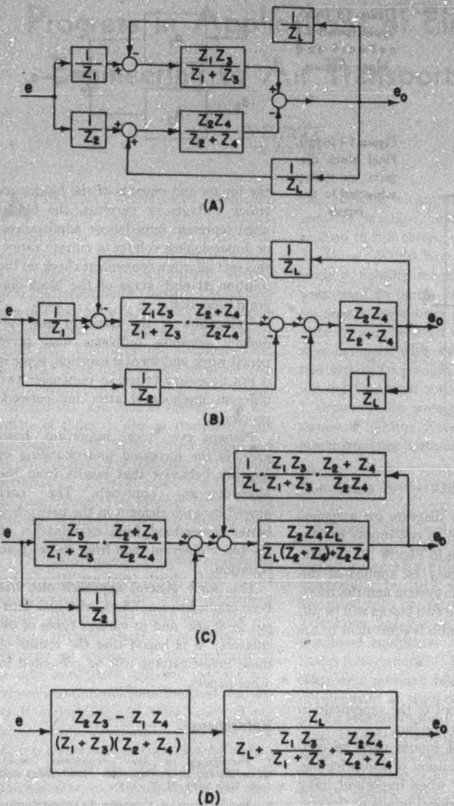

(A)
(B)
(C)
(D)
e
e_0
$\frac{1}{Z_L}$
$\frac{1}{Z_1}$
$\frac{1}{Z_2}$
$\frac{Z_1Z_3}{Z_1+Z_3}$
$\frac{Z_2Z_4}{Z_2+Z_4}$
$\frac{Z_1Z_3}{Z_1+Z_3}\cdot\frac{Z_2+Z_4}{Z_2Z_4}$
$\frac{1}{Z_L}\cdot\frac{Z_1Z_3}{Z_1+Z_3}\cdot\frac{Z_2+Z_4}{Z_2Z_4}$
$\frac{Z_3}{Z_1+Z_3}\cdot\frac{Z_2+Z_4}{Z_2Z_4}$
$\frac{Z_2Z_4Z_L}{Z_L(Z_2+Z_4)+Z_2Z_4}$
$\frac{Z_2Z_3-Z_1Z_4}{(Z_1+Z_3)(Z_2+Z_4)}$
$\frac{Z_L}{Z_L+\frac{Z_1Z_3}{Z_1+Z_3}+\frac{Z_2Z_4}{Z_2+Z_4}}$

Figure 11. Block diagram reduction for lattice network

the network. The usual conditions for zero output may be derived from the transfer function of the first block.

Typical bridged-T and parallel-T networks are shown in Figures 12 and 13. Reduction of the block diagrams, which requires procedures similar to those used for the lattice networks, is omitted to save space.

Even in their original form, the block diagrams show clearly the essential similarity of these networks. The fact that signals can arrive at the output by two different paths is the distinctive characteristic of these networks and the explanation of the null or balance properties which can be obtained in some cases. The required conditions for the equivalence of two such networks can be established by partial reduction of the block diagrams.

Transient and Frequency Response

At the beginning of the paper it was asserted that transient and frequency response are found with almost equal facility by the block-diagram method. A specific example, illustrating the general procedure, may serve to prove the point.

Since the transfer function is, by definition, the ratio of the Laplace transform of any output to the transform of the corresponding input, the block diagram supplies all of the data needed to find the response to any input. Thus, for the resistance-capacitance ladder network with two identical sections, we may write

$$e_2(s)=\frac{1}{R^2C^2s^2+3RCs+1}e(s) \qquad (19)$$

from equation 15. If, for example, the input is a suddenly applied constant voltage of magnitude E, then the transform is[5]

$$e(s)=\frac{E}{s} \qquad (20)$$

and therefore

$$e_2(s)=\frac{1}{R^2C^2s^2+3RCs+1}\times\frac{E}{s} \qquad (21)$$

Factoring the denominator and expanding in partial fractions,[5]

$$e_2(s)=E\left[\frac{1}{s}+\frac{0.171}{s+\frac{2.618}{RC}}-\frac{1.171}{s+\frac{0.382}{RC}}\right] \qquad (22)$$

The output as a function of time, by inverse transformation, is

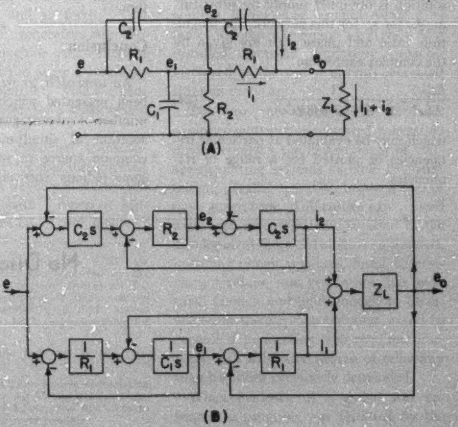

e_2
C_2
C_2
i_2
e
R_1
e_1
R_1
e_0
i_1
C_1
R_2
Z_L
i_1+i_2
(A)
C_2s
R_2
e_2
C_2s
i_2
e
Z_L
e_0
$\frac{1}{R_1}$
$\frac{1}{C_1s}$
e_1
$\frac{1}{R_1}$
i_1
(B)

Figure 12. Typical bridged-T network and block diagram

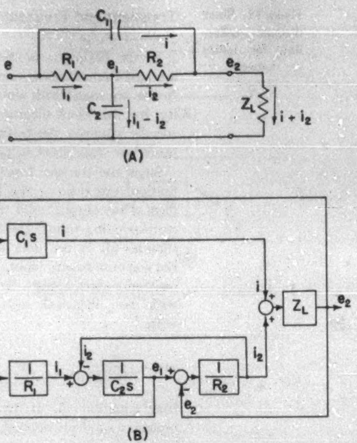


Figure 13 (left). Typical parallel-T network and block diagram

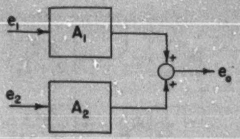


Figure 14 (right). Final block diagram for system subjected to two inputs

$$e_2(t)=E\left[1+0.171e^{-\frac{2.618t}{RC}}-1.171e^{-\frac{0.382t}{RC}}\right] \tag{23}$$

The response for any other input is determined in the same way, substituting the transform of the particular input and carrying out the partial fraction expansion and inverse transformation.

If the initial transient resulting from the application of a sinusoidal input is required, it may be obtained by use of the operational methods; the steady-state response will be obtained as part of the solution. If, on the other hand, the initial transient is of no concern and only the steady-state response is desired, the solution is obtained simply by substituting $j\omega$ for s; for this network, the magnitude ratio and phase shift are given by the complex expression

$$\frac{E_2(\omega)}{E(\omega)}=\frac{1}{1-R^2C^2\omega^2+j3RC\omega} \tag{24}$$

which may be evaluated at particular frequencies or plotted for a range of frequencies.

Multiple Inputs

The final block diagram for a system which is subjected to two inputs has the general form given in Figure 13. In general, the inputs may be applied at different points in the system and the transfer functions in the two blocks will be different. This diagram is equivalent to the equation

$$e_0(s)=A_1e_1(s)+A_2e_2(s) \tag{25}$$

which is a statement of the superposition principle for linear systems.

Any number of inputs may be considered by an extension of the principle. Initial conditions, when important, may be treated as additional inputs to the system.

Conclusion

An approach to network analysis has been presented which seems to have a number of advantages. By avoiding the solution of simultaneous equations, a common source of mistakes as well as some tedious work is eliminated. Since the inputs and outputs of the blocks are either voltages or currents, the blocks must represent impedances, admittances, or dimensionless voltage or current ratios; this fact affords a convenient check on the solution at each stage of the block-diagram reduction.

Unification of the transient and frequency response problems also saves pencil work and mental exertion, since it is not necessary to use or remember two different methods of attacking network problems.

Perhaps even more important, however, is the increased understanding of network behavior that results from the block-diagram approach. The part played by each element in the network is better understood from examination of the block diagram than from study of an equation.

This same general approach also has been applied to amplifiers, with and without feedback, and to various types of oscillators; it is hoped that the results of these investigations will be presented in other papers.

References

1. PRINCIPLES OF SERVOMECHANISMS (book), G. S. Brown, D. P. Campbell. John Wiley and Sons, New York, N. Y., 1948.

2. BLOCK DIAGRAM NETWORK TRANSFORMATION, T. D. Graybeal. *Electrical Engineering*, volume 70, November 1951, pages 985–990.

3. THE MULTISECTION RC FILTER NETWORK PROBLEM, L. Storch. *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 39, November, 1951, pages 1456–58.

4. BRIDGED-T AND PARALLEL-T NULL CIRCUITS FOR MEASUREMENTS AT RADIO FREQUENCIES, W. N. Tuttle. *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 28, January 1940, pages 23–29.

5. MODERN OPERATIONAL MATHEMATICS IN ENGINEERING (book), R. V. Churchill. McGraw-Hill Book Company, New York, N. Y., 1944.

No Discussion

Progress in Applications of Electrical Engineering to Air Transportation

D. E. FRITZ
MEMBER AIEE

APPLICATIONS of electrical engineering to air transportation encompass almost all the various phases of electrical engineering: lighting, communications, electronics, rotating machinery, switchgear, batteries, and so forth. The modern aircraft may have as many as 400 separate pieces of electric equipment, adding up to over 100 kilowatts of installed load; and may have three or four separate electric systems, with over 10 miles of feeder, distribution, and control wiring. The airplane of today is almost completely dependent on the electric power supply and is not capable of performing any mission without it. In most cases, a safe return to base, or any safe landing at all, is not possible without a ful'y functioning electric system.

The development of electric equipment for aircraft has been closely allied to the development of the main propulsion power plants. The chief difference between aircraft electric apparatus, and commercial apparatus is the emphasis on weight and space and the extreme environmental conditions imposed on aircraft apparatus. As more powerful engines were developed, airplanes flew faster, higher, and farther, thereby widening out environmental parameters and forcing new considerations on the electrical designer. Since many of the largest steps in engine development have been the result of military programs, resulting from need of a more powerful engine to make possible a new airplane that will outperform those of unfriendly nations, the periods of most intensive progress in electric aircraft equipment may be found during periods of war or military preparedness programs.

Early History, 1882–1917

The earliest recorded use of electricity on aircraft was the application of a storage battery and an electric motor to drive a propeller on a balloon in 1882. However, this did not prove to be a very successful means of providing propulsion power and was abandoned after a short trial period. At the beginning of practical powered flight, with the Wright Brothers airplanes in 1903, little or no thought was given to the airplane as anything other than a sporting machine, and the airplane remained in this category until World War I. During World War I, the airplane became a fighting machine and dirigibles were used to bomb London. Although no records have been found to substantiate it, it is believed that the dirigibles were equipped with some sort of an electric system to provide instrument lighting, and that some of the early World War I airplanes also were equipped with instrument lighting systems supplied with power from dry batteries.

The Years 1917–1930

Although radio installations had been made in aircraft prior to this time, the installations were not reliable and not really successful. In 1917, the Army, in cooperation with the Signal Corps and Bell Telephone Laboratories, installed radio telephone equipment in several United States Army aircraft. While the first systems were powered by storage batteries, development of wind-driven and engine-driven generators was initiated to provide more reliable power supplies. An electrical branch was established at McCook Field in Dayton, Ohio to develop lightweight electric equipment for aircraft, including generating systems, lighting, starters, batteries, and so forth. Two types of wind-driven generators were developed, one a d-c machine with a vacuum tube regulator, and the other a combination d-c generator and 900-cycle a-c generator to provide power for a spark transmitter.

When regularly scheduled transcontinental air mail service was inaugurated in 1924, equipment developed under the auspices of the electrical branch at McCook Field was available, and included engine-driven generators, controls, instrument lighting, landing lights, and radio equipment. Although the airplanes of the era were mostly equipped with hand-crank starters or were started by turning the propeller by hand, electric starters were developed to enable starting of the engines more quickly, and these in turn required that the storage batteries on the airplanes have larger capacity.

Many of the early aircraft electric systems were 8-volt systems, but 12-volt systems also were used and became an accepted standard for aircraft around 1925. The Boeing Airplane Company *Model 247* airplane, which was first produced for passenger service in 1931, was equipped with two 50-ampere 15-volt generators, which were totally enclosed, and also was furnished with electrically operated retractable landing gear. Inasmuch as the environmental conditions in which this equipment had to operate was not too much different from sea level operating conditions, the equipment was similar in design to automotive equipment of that period, except that considerably more output was obtained from a given size and weight. It was very unusual for airplanes to fly at altitudes above 8,000 feet, and blind flying was done only in extreme emergencies.[1]

The Years 1930–1940

The increased use of the airplane in passenger service resulted in much greater emphasis on passenger comfort items including lighting, ventilation, and schedule reliability. Electric loads rose sharply during this period, with the addition of new types of radio equipment, passenger reading lights, and ground ventilation fans. Reference to Figure 1 will show how installed electrical generating capacity increased as the use of aircraft increased through the years.

The use of multiengine airplanes equipped with engine-mounted generators made necessary the development of voltage-regulating systems capable of maintaining parallel operation between two or more generators. Difficulties experienced with system components resulted in the development of special aircraft equipment, including circuit breakers, fuses, reverse-current relays, and regulators that were small enough and light enough to be installed in the airplane but that would be capable of performing their required functions with a greater degree of reliability than had been previously demanded.

When evidence of a global war appeared, a program was initiated by the

---

Paper 52-266, recommended by the AIEE Air Transportation Committee and approved by the AIEE Committee on Technical Operations for presentation at the Centennial of Engineering—1952, Chicago, Ill., September 10–12, 1952. Manuscript submitted July 7, 1952; made available for printing August 22, 1952.

D. E. Fritz is with the Jack and Heintz, Inc., Cleveland, Ohio.

The author wishes to express sincere appreciation to those members of the Air Transportation Committee who contributed information and data to make the writing of this paper possible.

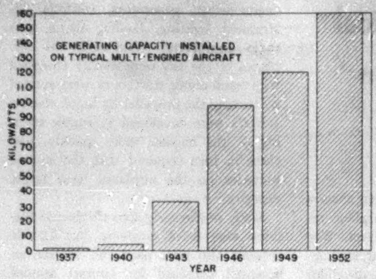


Figure 1. Generating capacity installed on typical multi-engine aircraft, 1937 to 1952

Armed Services to develop large bombers, and several aircraft companies initiated preliminary design studies of 4-engine airplanes designed to accomplish long-range bombing missions at higher altitudes and speeds than fighter airplanes of that time. Early in the study phase of these developments, it was realized that the long runs of very large wire, required to distribute the electric power through these airplanes with a 12-volt electric system, would result in untenable weight penalties.

Since the need for higher-voltage systems for large aircraft was apparent, consideration was given to higher d-c system voltages such as 24, 60, and 120 volts. As it was a relatively simple matter to convert 12-volt systems to 24-volt systems, immediate programs were initiated to accomplish this on most military aircraft. These programs were successful, and the 24-volt system became standard for all but the smallest aircraft.

The Navy had been using 800-cycle 120-volt single-phase a-c generators for many years to power radio receiving and transmitting equipment on their aircraft. This system had proved itself to be reliable and practical in actual service for radio power supplies, but many engineers believed that it would not be practical for a primary aircraft electric system, due to the necessity of using split-phase or capacitor motors and the large number of poles necessary to obtain reasonable motor speeds.

In 1935 the Air Force issued two design study contracts to evaluate the possibilities of using high-voltage alternating current on large aircraft. One of these design study contracts resulted in a recommendation to use an 800-cycle single-phase 120-volt a-c system, using split-phase capacitor motors, where motors were required. Development contracts for the necessary equipment were let to various manufacturers to provide a set of equipment for a large 4-engine airplane, which later was designated as the *XB-15* and was built by the Boeing Airplane Company. The generating equipment was driven by auxiliary engines and this equipment, together with other pieces of load equipment, was delivered to Wright Field in the summer of 1936 and installed in the airplane during the fall of that year.

The other design study resulted in a recommendation to produce 400-cycle 3-phase 208/120-volt systems, and development contracts for this equipment were let in parallel with the previously mentioned contracts for the 800-cycle equipment. The equipment was installed in an airplane built by the Douglas Aircraft Company and known as the *XB-19*. Both of the airplanes using these systems were flown for many hours on evaluation tests, and although many difficulties were experienced with the individual pieces of equipment, the results were sufficiently encouraging to arouse considerable interest in a-c systems for large aircraft. Considerable difficulty was experienced with the auxiliary engines used to drive the generators in both airplanes, and it is believed these difficulties resulted in decisions being made that the most reliable and efficient source of power to drive generators was from the main engine.

The choice of a standard system frequency for a-c systems was made only after a co-ordination meeting was held with the Air Force, Bureau of Aeronautics, British Air Ministry, and the electrical industry. A frequency of 360 cycles per second was proposed to enable a multiple of 60 to be used, but a standard frequency of 400 cycles per second was established.

Although the design studies and the tests on the *XB-15* and *XB-19* which were conducted showed that the 3-phase a-c systems had an appreciable weight advantage over the lower voltage d-c systems, an active controversy grew up among various airframe companies, government employees, and manufacturers as to which system would constitute the best system for a large airplane. Many papers were written on the subject and will be found in the technical literature of the day.[2]

The Years 1940–1945

Inasmuch as the main engines did not run at constant speed, and parallel operation of a-c generators was considered mandatory, development contracts were initiated by the Air Force to develop a variable-ratio transmission and suitable governing schemes to allow constant speed operation of the generators with sufficient accuracy to permit parallel operation. Other schemes of obtaining constant frequency were proposed, such as electronic frequency converters, but preliminary investigations and experiments did not show these to be practical for aircraft operation. One company was given a development contract to build an experimental version of a step-type planetary transmission and fluid drive, while several companies were given contracts to develop variable-ratio hydraulic drives.

About this time, the military established a program of developing intercontinental bombers. Contracts were let to Consolidated Vultee Aircraft Corporation to build the *XB-36* airplane, and to Northrop Aircraft, Inc. to build the *XB-35* airplane. Both of these airplanes were planned to use a constant frequency 400-cycle 3-phase 208/120-volt a-c system, and alternators driven by the main engines through constant-speed drives. The constant-speed drive that looked most promising was specified for both airplanes.

In order to provide power on the ground when the main engines were not in use and emergency power in the air, both airplanes were to be equipped with two 37.5-kva auxiliary power units. Development contracts were initiated both by the Air Force and by the airframe companies building the airplanes to provide accessory equipment necessary to make these airplanes almost completely a-c equipped, although a small quantity of direct current was to be provided for control operation such as operating of contactors. In one case, this direct current was to be provided by rotating converters rated at 200 amperes, and in the other case, transformer-rectifiers were used. Due to difficulties with the development of the constant-speed drive, both airplanes were initially flight tested with all power being supplied by the auxiliary power units. However, these difficulties were overcome, and although the *B-35* airplane program was cancelled, the *B-36* is flying today with a 3-phase a-c system. The operating experience obtained on the *B-36* airplane has been invaluable in the design of other a-c systems, and the experience gained has resulted in additional equipment being provided to make a more completely integrated system.

Although many difficulties were experienced with the a-c system and constant-speed drives on the early *B-36* airplanes, none of these were in any way due to the basic choice of system, but rather were due to mechanical difficulties in the equipment, or the use of compromise components that were not adequately designed for the intended service. Most of the difficulties were mechanical, and have been resolved to the point where both the constant-speed drive and the other various components are now considered satisfactory for military use. Major difficulties in obtaining competent maintenance personnel for the airplanes were experienced by the Air Force, due to a lack of trained personnel for servicing the 3-phase system. As the result of difficulties in paralleling and maintaining constant frequency, a much improved governor has been developed which enables very accurate paralleling and that will hold frequency to an accuracy of approximately 1/2 of 1 per cent.

As the result of the emphasis put on the *B-35* and *B-36* programs, the development of higher voltage d-c systems was hampered due to the lack of development capacity. However, the 120-volt d-c system was specified for use on the 8-engine Hughes flying boat *HK-1*, and a complete 120-volt d-c system was installed on this airplane.[3] Another large bomber was scheduled to use a 120-volt d-c system but was cancelled before an actual installation of the equipment had been made. One twin-engine fighter airplane has been built and put into production which uses the 120-volt d-c system as an auxiliary system to power fire control and radar equipment. The lack of completely developed utilization equipment for the 120-volt d-c system has resulted in its not being considered seriously as a primary system for new aircraft designs, while the 3-phase a-c system is gaining favor rapidly as a primary electric system for multiengine aircraft.

The 28-volt d-c system proved to be the workhorse throughout World War II. As previously mentioned, when the loads on single-engine airplanes became greater, the capabilities of the 12-volt system in use on some of the airplanes were exceeded, and they were changed over to 24-volt systems. This was relatively easy to accomplish because utilization equipment for 28-volt systems had been developed for other aircraft. On multiengine airplanes, 28-volt systems were in practically all military aircraft by 1944, and a complete line of generating and regulating equipment was available, ranging from 50 amperes to 300 amperes. As airplanes began to operate at higher altitudes, the limitations of fan cooling became apparent, and most aircraft were modified to provide ram-air cooling to the generating equipment. As system capacity increased, there was a tendency to make larger versions of existing equipment in an attempt to make it operate satisfactorily. This led to many difficulties inasmuch as the basic designs of such pieces of equipment as Terrill regulators were such that they had very definite limitations as to current capacity and, therefore, did not give satisfactory service life. Other types of regulators were designed and put into production, but ultimately were superseded by the carbon-pile regulator as used today.[4,5]

When additional electronic, radar, and radio equipment was added to the airplanes, a source of a-c power became necessary, and inverters, or motor-generator sets, were developed to provide 400-cycle a-c power which easily could be transformed and rectified to produce the high voltages for electronic tube plate circuit supplies, and low voltages for filament supplies. Since there were also several pieces of equipment on the airplane, such as flight instruments, that required 26-volt 3-phase 400-cycle inverters were used to supply this type of power.[6]

As military operations increased, additional pieces of radio and radar equipment were installed in aircraft, and the advisability of supplying the a-c power from inverters was questioned. Studies of the equipment utilizing the a-c power revealed that a majority of the load could be supplied by variable-frequency alternators which could be mounted on the main engines. Development contracts were initiated to develop variable-frequency alternators operating over a 400-cycle to 800-cycle range, thereby eliminating several inverters and reducing the load on the d-c system, which had become loaded to its capacity. Figure 2 shows a 20-kva generating system of recent design, with the electronic-trimmed carbon-pile regulator, panelized protective control, including ground and feeder fault protection, overvoltage protection, exciter field relay, and mounting tray.

When high-altitude daylight bombing missions were inaugurated over Europe, it was found that several electrical system failures occurred that were attributed to rapid brush wear at high altitude. A

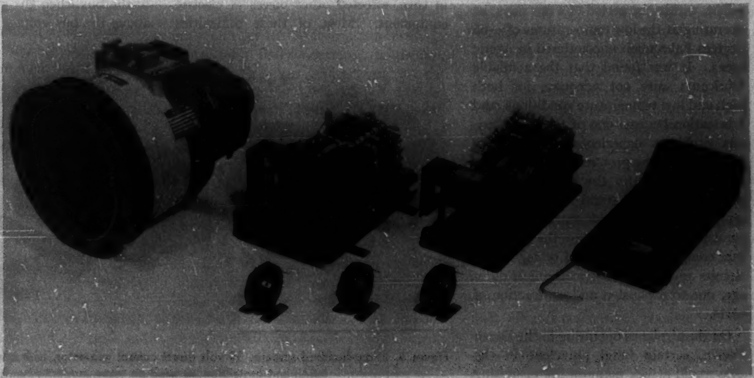


Figure 2. A typical set of power equipment for a 20-kva single-phase variable-frequency aircraft electric system

test program was initiated at Patterson Air Force Base by the Air Force, and it was found that at altitudes above 20,000 feet, brush wear was almost unpredictable, and that it was possible to wear out a complete set of generator or inverter brushes in a matter of minutes. Experiments on the part of brush manufacturers and equipment manufacturers resulted in establishing the theory that the lack of moisture in the air at high altitudes resulted in the loss of the copper oxide lubricating film on commutators and slip rings, thereby causing rapid abrasion of the brushes. Many different solutions to eliminate the rapid brush wear were tested, including wax impregnating of the brushes, introduction of engine oil fumes to the equipment, and separate brushes whose sole purpose was to apply lubrication to the commutator. However, none of these solutions were completely satisfactory and it was only with the introduction of halide-treated brushes as described by Dr. Howard Elsey, that operation at high altitudes without occasional unpredictable failures of brushes became possible.[7,8]

At the present time, with the improved treatments that are now available, operation of brushes at high altitudes does not seem to be a problem as long as sufficient cooling of the machinery is provided. However, there is evidence to indicate that as jet airplanes begin to fly for longer periods of time at high altitudes, the brush wear problem may again become serious.

As the United States Air Force and Navy began to operate airplanes all over the globe, difficulty was experienced with the wide extremes of environmental conditions that were encountered. An airplane operating one day in the desert at a temperature of +130 degrees Fahrenheit might, within the period of a few days, be operating at the low temperatures of −65 degrees Fahrenheit encountered in Arctic areas. It was found that the available lubricants were not adequate for both high and low temperature conditions, and lubrication became one of the more serious problems experienced on rotating machinery. The problem was solved, for the time being, by the development of greases that would stand wider operating temperatures. The wide extremes in temperature also would destroy the temperature compensation of tuned or regulating circuits and could cause mechanical damage, due to expansion and contraction of parts.

On the airplanes operating in the South Pacific, certain fungi prevalent in the area would attack many of the insulating materials and lubricating compounds used in electric machinery and cause failure of the equipment in a very short time. The high humidity prevalent in this area, together with salt spray in the air, caused corrosion problems and insulation deterioration to the extent that the failures of electric equipment seriously jeopardized the military program. Suitable fungicides and nonorganic insulation material were applied to the equipment, together with much better impregnating compounds and techniques, which resulted in acceptable solutions to these problems.

Mechanical failures due to vibration seemed to increase each time an engine or propeller modification was made. Apparatus that was operating satisfactorily on a given airplane might suddenly begin to fail after a few hours of operation if a different model of propeller was installed on the airplane or if the same equipment was applied to an airplane that had a different type engine. Equipment manufacturers were continually revising their designs to protect against vibration ambients that caused premature failures of their apparatus on airplanes. Due to the lack of sufficient information as to the actual vibration ambients, much of this work was done on a cut-and-try basis, and often modifications that were supposed to make a part stronger, actually resulted in earlier failure.[9]

Due to most of the components of the electric system being designed to objective specifications for the individual components, and the lack of adequate over-all system design concepts, many of the components were not capable of functioning suitably in the system to which they were applied. Many pieces of load equipment would reflect undesirable load characteristics back into the power supply which in turn, affected other pieces of electric equipment. Most of these difficulties were worked out on a cut-and-try basis by the aircraft manufacturers or service representatives in the theaters of operation. When the extent of this difficulty was realized, various organizations such as the Air Force, Navy, and several manufacturers of both airframes and equipment, began to study the over-all aircraft electric system and specifications were modified to result in components that were acceptable from a systems standpoint.

Experience on other phases of the electrical industry was drawn upon to provide solutions to many of the electrical engineering problems of aircraft. Although most of the same problems had been encountered in other industries, many of the solutions that had been used were not suitable solutions for aircraft, where weight and space were of prime importance.

Many new pieces of equipment were developed to provide a degree of system protection without penalizing the aircraft from a weight standpoint. The evolution of the "systems concept" did not come about in a short time but as a gradual process, because of detailed difficulties experienced with the components then in use. As an example of the new components that were developed to solve particular difficulties, the differential-voltage type of reverse-current cutout was the solution to a problem which caused automotive type of cutouts to fail in a few hours. The problem developed with the use of paralleled multigenerator systems where voltage regulator tolerances allowed the cutouts to operate and then be subjected to reverse current which caused them to open. This process repeated so rapidly it became known as "chatter." The differential-voltage reverse-current relay connects the generator to the bus when the generator voltage is above the bus voltage; therefore, no re-

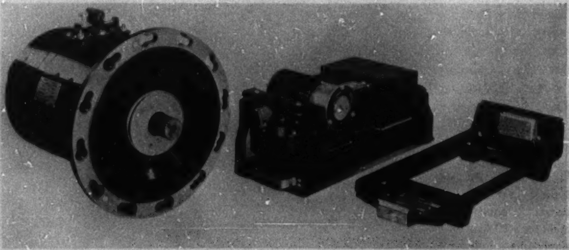


Figure 3. Four-hundred-ampere, 30-volt direct current generator, with control panel for regulation and protection, as used on many modern aircraft

verse current flows and the cutout remains closed once it operates.[10]

The addition of remote-operated armament on large bombers such as the *B-29*, resulted in a considerable increase in the electrical load on the system during combat maneuvers. In addition, the remotely operated armament resulted in a considerable increase in the electronic equipment installed in the airplane, together with several pieces of rotating machinery being added to supply the power for the armament turrets. Electrical loads were increased to the point where existing generators were no longer adequate, and it became necessary to install larger 300-ampere generators. In the meantime, the number of other electric devices utilizing electric motors was continually increasing to perform additional functions on the aircraft, which further increased the electric system load.

In spite of the emphasis that was put on high altitude operation, with practically all military specifications for electronic and electric equipment specifying continuous operating altitudes of 35,000 feet, very little actual operation during World War II took place above 25,000 feet. The aircraft speeds at high altitude were relatively low, and since most of the equipment that had large losses was cooled by ram air, cooling did not present a problem.

The use of jet engines in the latter phases of the war resulted in completely new environmental conditions for aircraft electric equipment. It was possible for these airplanes to fly at much higher altitudes than the piston engine airplanes, and they reached those altitudes in such short periods of time that the rapid change in pressure and temperature created a completely new set of problems. Airplanes making a rapid descent from the cold high altitude to a warm, humid sea-level condition would occasion moisture condensation inside electric and electronic equipment, and failures would occur due to arc-overs. Thus began a new period of aircraft operation which would create an entirely new set of operating conditions for electric apparatus.

The Years 1945–1952

Immediately after the cessation of hostilities of World War II, several airlines began negotiations for new transport airplanes. The experience and difficulties of World War II were foremost in the minds of operating personnel and airframe and electric equipment manufacturers. All recognized that if some of the more prevalent difficulties with electric equipment on military airplanes were to be avoided, the electric system for the new transport airplanes must be considered from an over-all system standpoint, rather than by considering individual components.

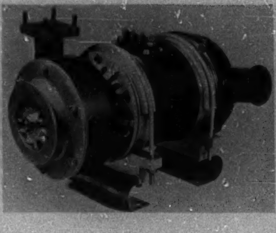


Figure 4. An air-turbine-driven alternator rated at 4 kva

The airframe companies developing the new transports had staffed their organizations with competent electrical engineers and electrical design sections. Operating in co-operation with the various electrical manufacturers, electric systems were designed which provided co-ordinated protection to the maximum degree allowed by the available equipment. Among these airplanes were the Douglas *DC-6*, Convair *240*, Boeing Stratocruiser, Martin *202* and the Lockheed Constellation. These airplanes had much higher electrical loads than previous transport airplanes and were equipped with 300-ampere generators on each engine. Additional loads, such as electrical deicing of windows, ventilation blowers, and blind flying equipment, were included as well. Although all of the airplanes were equipped with generating-system protection such as overvoltage; the Martin *202* was equipped with a completely packaged protective system, including ground-fault and feeder protection, overvoltage protection, reverse-current protection, and so forth. New military aircraft designs also were initiated and the same degree of application of the systems concept was applied to these designs, with several of them being designed for the use of completely integrated panelized electric systems.

The industry was now studying the co-ordination of various protective devices, together with characteristics of the equipment, in regard to transients, stability, fault-clearing ability, and fuse and circuit-breaker co-ordination characteristics. Since electric flight instruments had become standard, and electric engine controls were being used, continuity of electric power was more important than it ever had been before. Although most of the airplanes used in World War II could not perform their mission adequately if they suffered a loss of electric power, the airplanes being designed at this time probably would not be able to return to their home base without electric power. This was especially true since the airplanes were being designed for all weather operation and the blind flying and landing items were an absolute necessity. Extreme emphasis was, therefore, put on maintaining continuity of electric power. Figure 3 shows a 400-ampere 30-volt direct-current generator, together with a panelized regulating and protective control system and its mounting tray. The control system was developed to an Air Force-Navy-Industry co-ordinated specification, and includes ground-fault protection, overvoltage protection, reverse-current protection, paralleling provisions, as well as voltage regulation.

The development of the jet engine had progressed rapidly and power plants were available to produce many times the horsepower previously available from piston engines. The aerodynamic efficiency of airplanes designed to use jet engines resulted in much lower drag and much higher air speeds to the point where a definite change in environmental conditions was evident. Airplanes began to operate for considerable periods of time at altitudes in excess of 35,000 feet, and at flight speeds fast enough to result in an adiabatic rise in the temperature of the air. The low air temperature at altitude no longer compensated for the decrease in efficiency of cooling fans for electric and electronic equipment, and in addition, the engines on jet airplanes operated at relatively high speeds at all times when the aircraft was under powered flight.

The high wing loading of present-day jet airplanes results in a much greater penalty, if a pound of additional weight is put into equipment, than on previous slower-flying aircraft. In addition, the effect of a small amount of drag is multiplied greatly in the over-all performance of the airplane. It is, therefore, very important that the weight of electric equipment installed in jet airplanes be as light as possible, and that all possible surface drag be minimized. The small frontal area of axial-flow jet engines, together with the increasing size and capacity of engine-mounted equipment, has resulted in a review of the practicability of removing the electric generators from the main aircraft engines. A possible way of accomplishing this is to use a hydraulic drive wherein the hydraulic pump is installed on the engine, and a hydraulic

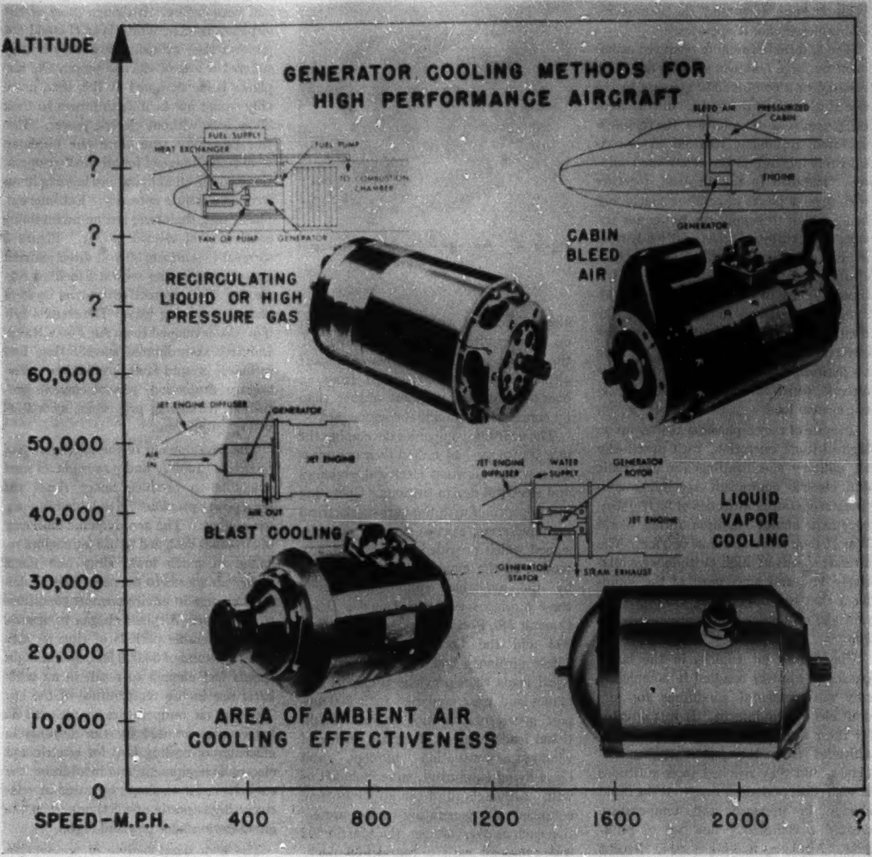


Figure 5. Methods of cooling aircraft generators applied to high-performance aircraft, with typical machines of current design

motor driving the accessories is located in another part of the airplane.

Another possibility is the use of compressed air taken from the compressor stages of the jet engine and piped through an air turbine which drives the accessories and is mounted remotely from the engine. This system has gained favor within the past few years, and several airplanes soon will be flying utilizing this system. It offers several advantages but also has the disadvantage of bleeding air from the compressor section of the jet engine, which under high-altitude operating conditions may seriously impair the performance of the engine. However, it appears that for certain applications this may be the most practical way of obtaining accessory power from the main engine, without mounting the accessories directly on the engine. Figure 4 shows a 4-kva air-turbine-driven alternator that is typical of this type of power supply. Auxiliary power units also are being considered for large aircraft, but instead of using piston engines, gas turbine designs are being developed.

The constant-speed drive has been developed to the point where it is considered a reliable device and is much more accurate in frequency control and paralleling than the original units. Constant-speed drives are now being built into the engine as part of the engine, instead of being bolted to the engine accessory case. Serious consideration is being given to the incorporation of a large alternator inside the case of the jet engine. However, there are several problems to be overcome on multiengine airplanes before this is accomplished, since most airframe designers still believe that it is necessary that two or more alternators be operated in parallel, to provide the necessary degree of reliability that is needed. Serious consideration is being given to dividing the various aircraft loads so that parallel operation will not be required, and accept-

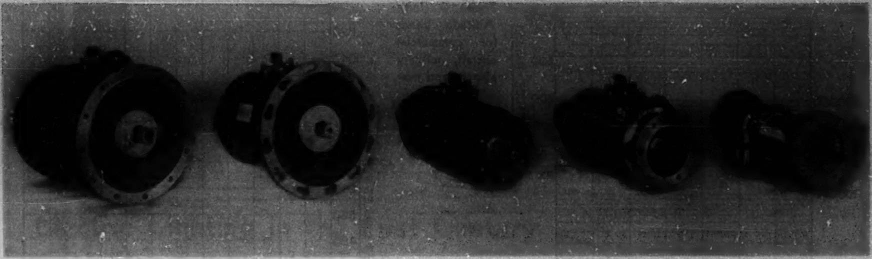


Figure 6. Typical d-c aircraft generators of current design, ranging from a rating of 35 amperes, 30 volts on the right to 500 amperes, 30 volts on the left

able reliability will be obtained by transferring loads, in the case of a failure of one generator or one engine.

One of the most significant developments of the present era is the emphasis being given to the study of the entire electric system of the airplane, together with the load utilization devices, from an over-all system standpoint. In many cases, the problem of getting rid of the heat generated by the losses of the electric machines requires that the design of the electric machine be integrated with the design of the airplane, to avoid serious weight and performance penalties on the airplane. As an example, on many of the high-speed high-altitude airplanes, air taken in by an air scoop on the outside of the airplane will not serve to cool the electric equipment, but instead, the air may be hot enough to result in damage to the equipment. It is, therefore, evident that the heat must be rejected in some other manner. Methods of accomplishing heat rejection from electric generating equipment include the following possible methods: the use of evaporants circulated through the machine, liquid or gas circulated through the machine, rejecting heat into the aircraft fuel supply through a heat exchanger, and the refrigeration of air taken in from outside the cabin, or from the jet engine compressor stages. Illustrations of these possible methods will be seen in Figure 5. Figures 6 and 7 show a-c and d-c generators typical of generators used in present-day aircraft and large missiles.

The space limitations of the modern jet airplane are much more stringent than those of previous airplanes, and every effort to miniaturize the equipment must be made if it is to be installed in the available space in the airplanes. The problems involved in rejecting the heat of the equipment become more serious as the equipment is made smaller and more compact because, in general, the efficiency does not increase when the item is miniaturized. In the case of many electronic components, miniaturization programs have resulted in pieces of equipment, occupying the cubical content of one-tenth of the space occupied by previous designs.

Almost all of the protective devices used on utility systems have equivalent items on the electric generating system of the modern aircraft. Examples of generating-protecting devices for a-c and d-c systems are shown in the following

A-C Protective Devices

Ground fault relays
Overvoltage relays
Over and under frequency relays
Reverse power relays
Phase sequence relays
Exciter ceiling protection relays
Paralleling provisions
Automatic synchronizing control

D-C Protective Devices

Ground fault protection
Overvoltage protection
Reverse current protection
Paralleling provisions

The development of guided missiles has also resulted in a new class of aircraft electric apparatus. Missiles usually have relatively short-time flight patterns, and extremes of altitudes up to 200 miles have been reached with missiles during the recent years. Speeds are so high that the surface of the missile can reach dull-red temperatures, and the problem is one of preventing the heat generated by the missile from reaching the equipment, and at the same time rejecting heat from the apparatus.

Shock and vibration characteristics of missiles are much more severe than those encountered in normal aircraft operation. The problems of miniaturization of equip-

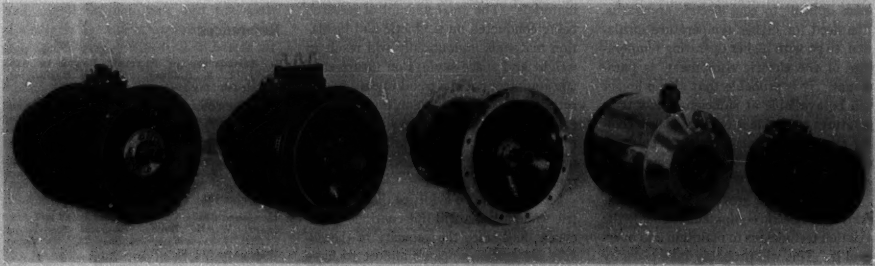


Figure 7. Typical a-c aircraft generators of current design ranging from 2.5 kva on the right, to 30 kva on the left

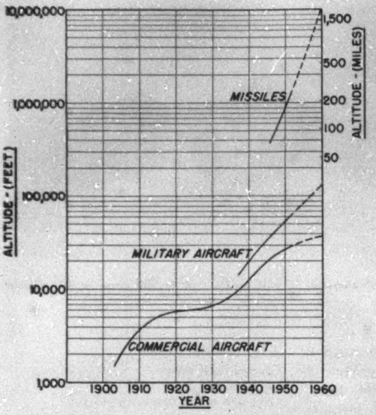


Figure 8 (left). Altitude trends for commercial aircraft, military aircraft, and missiles

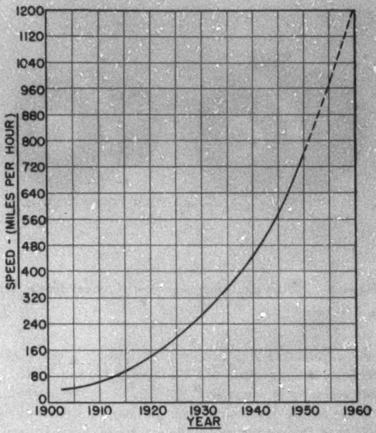


Figure 9 (right). Flight speed trend of piloted aircraft

ment are even more stringent in missile equipment, due to the fact that it is necessary to pack as much fuel as possible into the missile, and very little space is left for electronic and electric equipment. Since most missiles are used but once, the life requirements of missile equipment are different from those for equipment used on conventional aircraft, even though long life is not necessary. Extreme reliability is a necessity because the failure of a single component can result in the loss of a very expensive missile. The power supplies presently used on most missiles have been batteries, but the development of small turbine-driven generators is progressing.

Trends for Future Development

It is anticipated that the future will bring more serious consideration of the electric equipment as a part of the over-all airplane. As more-powerful engines are developed, and airplanes fly at higher altitudes and high speeds, it will become more difficult to cool electric apparatus. The need for higher temperature insulation, to go with higher operating temperature of the electric equipment, is paramount. It is probable that air no longer will be available at sufficiently low temperatures to enable it to be used as a cooling agent, and that completely environment-free operation of electric equipment will be needed.

There is also a great need for simplification of electric equipment to enable the aviation technicians to maintain and overhaul it. This will probably take place by sectionalizing components to the point where complete circuits are removed and replaced as plug-in units. New developments of recent years, such as transistors and germanium diodes, are expected to simplify and reduce the size and weight of electronic and control devices. The complexity of the electric system and equipment on the modern fighter airplane must be reduced if the airplane is to be used as an effective weapon. In order to accomplish this objective, the designers of the equipment must study the airplane as a system, and must avoid as many pieces of conversion equipment as possible. Each protective device or circuit, must be evaluated as to the worth of the protection, versus the additional hazard it creates. The probability factors of the various types of failures must be recognized and used as the criteria to decide if additional weight and complexity of protective devices and standby circuits and equipment are justified.

The designer of aircraft electric equipment must always take advantage of the special characteristics of new and special materials. The great amount of research being conducted on magnetic and insulation materials undoubtedly will result in weight and space savings on future electric equipment. The trend toward higher rotational speeds is evident, and electrical engineers will devise ways of building rotating machinery that will give satisfactory operating life at speeds of 30,000 rpm and higher.

Since there is no evidence that the electric system loads will not continue to increase, and since the capacity of the 28-volt d-c system appears almost to have been reached, it appears that most future large aircraft designs will use the 400-cycle 3-phase a-c system as the primary system, with other systems being used as auxiliaries.

Future aircraft and missiles that operate at much higher altitudes and flight speeds will cause drastic changes in the presently accepted environmental standards for aircraft electric equipment. Some of these new environmental parameters will be as serious as the high-altitude brush wear problem in World War II. Figures 8 and 9 will show the trend of altitude and flight speeds from which one may obtain an indication of future environmental requirements.

It is certain that designing electric equipment for use in aircraft will continue to be one of the most challenging fields for electrical engineers, as long as there is need for improvement in the performance of the aircraft. Until that period is reached, standardization and cost will continue to be secondary to many other factors, including weight, space, and reliability.

References

1. APPLICATION OF ELECTRIC POWER IN AIRCRAFT, T. B. Holliday. *Electrical Engineering*, volume 60, May 1941, pages 218–25.

2. BASIC CONSIDERATIONS IN SELECTION OF ELECTRIC SYSTEMS FOR LARGE AIRCRAFT, W. K. Boice, L. G. Levoy. *Electrical Engineering* (*AIEE Transactions*), volume 63, June 1944, pages 279–87.

3. HIGHER-VOLTAGE D-C AIRCRAFT ELECTRIC SYSTEMS, W. L. Berry, J. P. Dallas. *Electrical Engineering* (*AIEE Transactions*), volume 63, November 1944, pages 843–49.

4. CARBON PILE VOLTAGE REGULATORS FOR AIRCRAFT, W. G. Neild. *Electrical Engineering* (*AIEE Transactions*), volume 63, November 1944, pages 839–43.

5. AIRCRAFT CARBON-PILE VOLTAGE REGULATORS, FUNDAMENTALS, AND DESIGN IMPROVEMENTS, B. O. Austin, H. H. C. Richards. *AIEE*

Transactions, volume 68, part II, 1949, pages 983–87.

6. 400-Cycle Inverters for Military Aircraft, C. P. Hayes, L. L. Ray. *Electrical Engineering (AIEE Transactions)*, volume 64, May 1945, pages 233–37.

7. Treatment of High-Altitude Brushes by Application of Metallic Halides, Howard M. Elsey. *Electrical Engineering (AIEE Transactions)*, volume 64, August 1945, pages 576–79.

8. High-Altitude Brush Problem, D. Ramadanoff, S. W. Glass. *Electrical Engineering (AIEE Transactions)*, volume 63, November 1944, pages 825–30.

9. Aircraft-Electric-Accessory-Vibration Investigations, D. R. Miller. *AIEE Transactions*, volume 63, 1944, pages 1283–90.

10. An Aircraft Differential-Voltage Cut-out, O. C. Walley. *Electrical Engineering (AIEE Transactions)*, volume 63, September 1944, pages 632–34.

No Discussion

Electric Lighting in the First Century of Engineering

R. L. OETTING
MEMBER AIEE

THIS centennial year is a fitting time to set down a few of the historical facts regarding the men and the events that created and expanded the production and application of electric light as a science, an art, and an industry. Here, too, are recorded observations on the impact electric lighting has had upon our lives.

1852–1877 Early Investigations

By the middle of the nineteenth century, the conversion of electric energy into light had been investigated by a number of individuals. As early as 1802, Sir Humphry Davy of England demonstrated that light came from materials which were raised to high temperature by passing an electric current through them. He also discovered that light could be produced by causing an electric current to span a gap between charcoal electrodes. An elaborate demonstration employing a battery of 2,000 cells was presented in 1808 before the members of the Royal Institution. This was Davy's description of his experiment: "When horizontal pieces of charcoal, about an inch long and one-sixth of an inch in diameter, were brought near each other, within a thirtieth or fortieth part of an inch, a bright spark was produced, and more than half the volume of charcoal became ignited to whiteness; and by withdrawing the points from each other a constant discharge took place through the heated air, in a space equal to at least four inches, producing a most brilliant ascending arch of light." Had the charcoal pieces been arranged one above the other, sources of this type might not have been named "arc" lights by Davy in 1821. Davy went on to discover that a magnet would deflect the path of the arc. In this regard, the arc was similar to a wire carrying electric current.

Archereau in France was the first to feed the arc carbons with electromagnetic devices. This was about 40 years after Davy's demonstration. It was the original use of a principle still employed; there have been many other instances where associated electric equipment was adopted to improve the production and utilization of light. In 1856 J. Lacassange and R. Thiers invented the "differential" method of arc control. This helped make carbon arc lamps of some commercial value a few years later.

The distinctive blue-green color of mercury arcs was first seen at about this time, too. It was in 1860 that J. T. Way made the discovery by breaking a stream of mercury while it was connected to a battery.

De la Rue is credited with making a filament incandescent lamp with a coiled platinum wire in 1820. The glass tube was sealed with brass caps and supposedly evacuated. The first lamp patent was granted to Frederick DeMoleyns by the British government in 1841. It was a "burner" consisting of two coils of platinum wire containing powdered charcoal and operating in a vacuum. The first American to receive lamp patents was J. W. Starr who, working in England with financial support from George Peabody, used a slender solid rod of carbon in the vacuum above a mercury column. His patents were granted in 1845. Shortly afterwards, Sir Joseph W. Swan became interested in incandescent lamps. From 1848 to 1860 he made a number of lamps using carbonized strips and spirals of paper and cardboard. Lacking effective vacuum equipment, he discontinued his studies.

One of the earliest applications of electric lighting was in 1859 when Professor Moses G. Farmer lighted his home with carbon-rod incandescent lamps. He designed and made the lamps and the battery on which they operated. Another achievement was the construction of a miniature electric railroad on his farm in Maine. It established him as one of the earliest model railroad enthusiasts.

The Alliance Company of Paris made some of the first commercial installations of arc lights. In 1862, a dynamo and an arc were placed in the Dungeness Lighthouse off the English coast. V. L. M. Serrin designed a mechanism which moved both carbons. They were held by chains wound around drums which, in turn, were controlled by ratchets actuated by electromagnets. About 1866, Ladd tried to introduce arc street lights into London, but the citizens preferred the gentle glow of the gas lights. About 1872, however, the clock tower of Parliament was lighted by an arc whenever the Houses were in session. An early dynamo was used. Arc mechanisms were improved in 1876 by Lontin who incorporated series and shunt magnets for better control.

At the close of the American Civil War, a 16-year-old lad in Cleveland was busily trying for himself all the electrical experiments he found in the reference material of that time. He was so taken by the account of Davy's arc-light demonstration that he scavenged enough junk metal to build a battery and reproduce the experiment. This was Charles F. Brush who sold his first "commercial" dynamo and arc light to a Dr. Longworth of Cincinnati some 12 years later, in 1877. Brush thereby established a position of leadership in electric-arc lighting which he maintained for some years.

Although a gradual improvement in

Paper 52-267, recommended by the AIEE Production and Application of Light Committee and approved by the AIEE Committee on Technical Operations for presentation at the Centennial of Engineering—1952, Chicago, Ill., September 10–12, 1952. Manuscript submitted June 30, 1952; made available for printing August 11, 1952.

R. L. Oetting is with the General Electric Company, Cleveland, Ohio.

batteries enabled these early investigators to make some advances, the real opportunities and progress came with the introduction of effective mechanical generators. About 1866–67, Dr. Henry Wilde, Professor Farmer and C. F. and S. G. Varley independently developed self-excited, electromagnetic machines. Even earlier Soren Hjorth took out a patent on such a device. The first Gramme machine was shown in 1871 and became commercially practicable about 1873. No longer was the output of mechanical generators limited to the feeble magnetic fields of permanent magnets and no longer was light source development dependent upon battery power. No immediate attention could be given to application engineering, however, since there were still the problems of light production to be solved. Even the Welsbach gas mantle was some years away.

Among the investigators of this period were theorists and experimentors. One of the former was the Scottish James Clerk Maxwell whose fundamental concepts and equations were to influence all sorts of engineering applications for years to come. In 1856 he promulgated his theory of wave motion for light and, in 1873, extended it to introduce the electromagnetic aspect of light.

Two years later, a very practical man, Thomas A. Edison, sold his invention of a quotation ticker for $40,000 and established a workshop and laboratory in Newark, N. J. He moved it to Menlo Park, N. J., in 1876. Meanwhile, George Westinghouse was becoming established in the design and manufacture of railroad equipment in Pittsburgh. Elihu Thomson was lecturing at the Franklin Institute and teaching with his senior colleague Edwin J. Houston in Philadelphia's Central High School. Mutual interest in electric lighting and individual abilities were to make Edison, Westinghouse, and Thomson leaders but fierce competitors in the next period of lighting progress

1878–1902 Pioneers in Practical Invention

This might well be termed the era of individual recognition. Progress was rapid, even tumultuous at times, and discoveries and inventions were closely coupled with individuals. In this period, manufacturing and electric service companies were formed, absorbed, and consolidated. Expanding research and engineering development programs called for pooling of financial assets while consolidation was also an expedient to avoid strangling legal entanglements. These corporate developments became an essential link in the chain of lighting progress.

Four aspects of progress have been separately treated in this period:

1. The development of light sources.
2. The applications to which the sources were put.
3. The development of associated electric equipment.
4. The emergence of engineering and manufacturing techniques and facilities.

ARC LIGHTS

Arc sources were improved by a number of engineers, or electricians as they were then known. In 1876 Edward Weston designed a d-c arc with one carbon above the other and a differentially controlling mechanism. Arcs at this time consumed the upper carbon, which produced 85 per cent of the light, at 1 to 2 inches per hour. In 1877 cored carbons were introduced. A soft carbon core was placed within a hard outer shell to give a more practical balance between light output and rate of carbon consumption. About this time, Paul Jablochkoff, a Russian army engineer, developed the "electric" candle. Two carbon rods were placed side by side and insulated from each other with kaolin. The exposed ends were bridged by fine carbon to start the arc, but if it went out it could not be started again. Alternating current was used to obtain equal burning of the carbons.

In 1877 Brush developed 8-hour carbons and all-night street lighting was made possible. Two sets of carbons were installed in an arc lamp, the second set going into operation after the first was consumed. Brush also developed a regulating coil to provide uniform feed of the carbons and a shunt coil for series-circuit lamps. The latter gave continuity to the circuit when an individual lamp failed. Arc lighting could then be operated from a centralized source of power and the idea of central stations was born.

Elihu Thomson was the first to foresee that the use of a series transformer for each lamp on an a-c system would also assure the continuity of the circuit when a lamp failed. He demonstrated the principle with small arc lamps during his Franklin Institute lectures in the winter of 1878–79. However, his first order was for a d-c dynamo. As soon as it was designed, built, and successfully demonstrated, he immediately turned his attention to the design and production of arc lamps so that his company could furnish complete lighting systems.

Other improvements were made as the years went by. Beginning in 1886, William Jandus of Cleveland took out a series of patents on enclosed arc chambers. Another who found that restricting the amount of air fed to an arc prolonged the life of carbons was Louis B. Marks, later the first president of the Illuminating Engineering Society. He, too, received a number of patents starting in 1894. Enclosed arcs were quickly adopted in street lighting. In 1898 Hugo Bremer of Germany and Andrew Blondel of France independently discovered how to apply metallic salts to the carbons which increased efficiency enormously. Flame arcs of this type were introduced commerically in Europe and the United States. In 1900 Dr. Charles P. Steinmetz pioneered another major improvement. He produced electrodes of magnetite and titanium which were so brilliant that they were named "luminous arcs." The development was carried further by Thomas Dempster.

GASEOUS DISCHARGE LAMPS

There was also progress with electric discharge arcs. Heinrich Geissler of Germany had found in 1856 that a tube containing air at reduced pressure would glow in a high-voltage a-c field. Three years later, A. E. Bequerel placed certain powdered phosphorescent materials in such tubes and found that some also fluoresced. Crookes demonstrated his improvements in Geissler-type tubes in 1879, and Edison patented a Crookes tube coated with fluorescent calcium tungstate in 1896. Demonstrations of discharge-type sources at lectures by Nicoli Tesla in 1891 led D. McFarlan Moore to organize two companies bearing his name to develop larger, cooler, and whiter sources of light than the filament lamps then in use. He demonstrated 7- to 9-foot long tubes in 1895 and, by 1904, solved a "short-life" problem by automatically feeding more gas into the tubes as they operated. He first used nitrogen, then carbon dioxide, later rarefied air, and finally rarefied air freed of oxygen.

Sir William Ramsay alone and in cooperation with Baron Rayleigh and M. W. Travers succeeded in isolating the five rare and inert gases of the atmosphere. Argon was discovered in 1893 and neon, helium, krypton, and xenon in 1898. When used to fill Geissler-Crookes tubes, each gave a distinctive color of light. Georges Claude of France developed an economical method of producing neon in 1902 and pointed out the use of neon for lighting about 1910.

Arons made a mercury arc of limited use in 1892. Peter Cooper Hewitt developed the first practical mercury source in 1901. His lamp consisted of a straight tube with electrodes at each end. When the lamp was tilted, a column of mercury bridged the gap between electrodes. Upon release, the mercury column broke as it ran back into a reservoir and the electric discharge was started. Later, sparking, heating, and electromagnetic tilting devices were used and other shapes of tubes were introduced.

INCANDESCENT LAMPS

Early experiments established the fact that practical incandescent sources would have to be free of air. Some experimenters worked in the direction of vacuum lamps. Available vacuum equipment was not particularly effective, however, and progress was slow until Herman Sprengel invented his mercury vacuum pump in 1865. Sir William Crookes used it in 1875 to exhaust his famed radiometer. Swan resumed his incandescent lamp studies in 1877 with a young bank clerk, Charles H. Stearn, who was familiar with Sprengel's invention. They were among the first to learn that the vacuum in a lamp deteriorated as the lamp was used. Concluding that gases were evolved from the carbon filament, they overcame the difficulty by energizing the filament during the exhaust process.

Lamps made by William E. Sawyer in America were typical of most earlier attempts in that they employed graphite-rod burners. However, he used an atmosphere of nitrogen within the glass bulb which was fitted to a metal holder. The lamps were designed so they could be renewed with fresh burners.

Thomas Alva Edison first turned his attention to the study of incandescent lamps in the fall of 1877. He lighted his first practical model just 2 years later, in October 21, 1879, and in exactly 2 more months made a public demonstration of 60 lamps on the laboratory grounds, in several houses, and along some of the country roads in the neighborhood of Menlo Park, N. J. In contrast to the years of effort on the part of others, Edison's rapid progress, as well as the practicability of his work, made his contribution all the more impressive.

Patent number 223,898 was granted to Edison on January 27, 1880. It covered a lamp having this combination of features:

1. A filament of high, not low, resistance.
2. A high vacuum.
3. An all-glass enclosure.
4. All joints sealed by fusion of the glass.
5. Platinum wires imbedded in the glass to carry current to the filament.

The high-resistance filament was the result of Edison's analysis that multiple and not series circuits were the most suitable for incandescent lamps. Not all of the other features were original with Edison but the combination was and the refinements he introduced established him as the inventor of practical incandescent lighting. Edison's concern for a complete and economical system of electric lighting, in addition to a workable light source, set him apart from the other pioneers. His practical solution of these self-established objectives created the electric lighting industry.

The filament is the heart of an incandescent lamp. That in Edison's first lamp was carbonized thread but it and carbonized-paper filaments were porous and fragile. Searching for other easily carbonized materials, Edison tried fibers from a bamboo fan. The filaments were hard and strong. Men were sent to all parts of the world in the hopes of finding an even better variety of bamboo but none proved to be superior to that in the bamboo fan. It was one of the epic searches of American industry for a better material. In the years that followed, many another search seemingly failed, but the continuing re-search did not. Progress has been built upon scientific research.

The early "A" lamps for use on 110-volt circuits had an average horizontal candle-power of about 16 and a life of 600 hours at 110 volts. They were rated as "8 to the horsepower" or about 93 watts. "B" lamps for two-in-series operation were rated at 8 candlepower, 55 volts, and 600 hours. Late in 1881 the efficiency was improved so that 16-candlepower lamps rated 10 to the horsepower without a reduction in life. Individual lamps departed from these nominal values because of variations in filament material, length, and diameter. There was a wide spread in the resistance of lamps and the differences in light output were even greater.

In 1878, H. S. Maxim (inventor of the Maxim gun) developed a carbon-rod lamp with a rarefied hydro-carbon-vapor atmosphere. Later, Sawyer and Man found that heating a filament in gasoline vapor deposited graphitic carbon on it. The amount deposited could be controlled, and the uniformity of filament resistance was thereby improved. Thomson obtained a license to use the process in 1885 and had equipment developed to "flash" carbon filament lamps. By 1893, Leigh S. Powell of England had developed a machine which automatically gave this type of treatment. The operator had only to load filaments and remove them—and no longer sat next to a 2-gallon jug of gasoline. Mechanization reduced the hazard and cost of lamp making.

Early in 1880, following Edison's invention of a successful incandescent lamp, Swan learned that cotton thread treated with sulphuric acid lost its fiberous condition. It could then be bent into spirals which retained their shape during carbonization. Some years later, he and Powell separately developed squirted carbon filaments. Swan dissolved explosive nitrocellulose in acetic acid and squirted the filaments into an alcohol bath while Powell dissolved ordinary cotton in hot zinc chloride. The filaments set into strong, smooth, structureless cellulose threads which were available in long lengths and could be carbonized readily. Weston, Edison, and others in this country produced similar results with a variety of materials and processes. The control of filament resistance was improved and filaments of this type were widely used by 1894.

The Nernst lamp of 1897 was inspired by the Welsbach gas mantle. Developed by Dr. Walter Nernst of Germany, it was a marvel of ingenuity. Platinum-wire coils heated magnesia-rod "glowers" until they became conductors and light emitters. The coils were then disconnected by a magnet. Glowers were later made of zirconia and yttria and finally of ceria, thoria, and zirconia materials to become familiar in gas mantles and electronic tubes. Having a "negative" resistance characteristic, Nernst lamps required a current-limiting device. Ballasts of iron-wire coils in a hydrogen atmosphere were developed here.

During this period metallic filaments were becoming of interest. In 1898 Welsbach applied for a patent on a filament made by squirting osmium into strands which were then sintered. Osmium melted at a lower temperature than carbon but didn't vaporize as readily so that such filaments could be operated at a higher temperature. Osmium lamps of 5.9 lumens per watt were sold in Europe from 1899 to 1906.

Tungsten was of academic interest. The metal had been isolated by Scheele in 1870 who found it hard, brittle, and twice as heavy as lead. During 1893–94, Alexander D. Lodyguine, a Russian, attempted to plate tungsten around a carbon core. At the turn of the century, Alexander Just and Franz Hanaman were laboratory assistants to the chemistry professor of the Vienna Technical High School. They were just developing an interest in tungsten, an interest which finally led to the most practical filaments ever developed.

In the latter years of this period, references report the efficiency of most lamps as 3.5 to 4.5 watts per candlepower. Life designs were aimed at 1,000 to 2,000 hours although it was not infrequent to find lamps lasting 4,000 to 5,000 hours. Candlepower ratings of 8, 16, and 32 were the most popular, but appreciably higher ones were used occasionally. Well before the end of the nineteenth century, imitation of other sources developed. By 1892, candle-shaped lamps of 5 and 10 candlepower were advertised in England. Lamps with clear, one-half obscured, and tinted bulbs were made. Where source reliability was extremely desirable, such as the navigation lights on ships, double-filament lamps were supplied.

LAMP AND LIGHTING APPLICATIONS

Hayes had just followed Grant as President of the United States when the first electric-arc street lighting was installed in this country. Dr. Weston made the installation in Newark, N. J., during 1877. Jablochkoff's electric candles were used the following year to light the Place and l'Avenue de l'Opera in Paris as part of the Universal Exposition. It was a notable installation even though short-lived carbons and other limitations kept it from being sufficiently practical for widespread use. Jablochkoff developed the arcs which were self-regulating, an a-c generator, and a series system of energy distribution. A Professor Sillinan wrote from Paris in July 1878 about "the splendor of the electric illumination. The effect is magnificent and at this moment there exists nothing in this city of splendid effects to compare with this magical scene. The vista is about 2/3 of a mile and the effect incomparably finer than any show of artificial illumination ever before seen."

About the time the professor was writing this, Brush was installing twenty arc lamps in the Philadelphia store of John Wanamaker. Five dynamos were used—each machine having four armatures, and each armature supplying one lamp. Brush units using his 8-hour carbons were first sold in 1879; twelve were put into service in Cleveland on April 29, 1879, and sixteen were installed to illuminate Niagara Falls on July 4.

Thomson's first commercial equipment was installed in Fuller's Bakery in Philadelphia in 1879, the same year that an amusement park in San Francisco was lighted electrically. In 1880, Brush arcs were mounted on high towers erected in New York City's Madison Square. An installation in Wabash, Ind., was similar. Four arc lights were mounted on poles above the courthouse dome to shine into every street and yard "with a strange, weird light, exceeded in power only by the sun." By 1897 special arc lamps were developed for use on electric trolley circuits. The trolley companies owned or leased many outdoor amusement parks; electric lighting brought out the crowds.

The year 1880 marked the first commercial installation of Edison's incandescent lamps. The steamship *S. S. Columbia* was being built in Chester, Pa., for the Oregon Railway and Navigation Company. After launching, it was brought to New York where the Edison Electric Lighting Company installed 115 lamps. The second installation of Edison lamps was made in 1881. This was at 229 Pearl Street, New York City, for the lithographers Hinds, Ketchum and Company.

During 1881 and 1882 more than 30,000 lamps were installed in about 150 locations on steamships, in machine and car shops, mills, stores, offices, theaters, hotels, and homes. By the turn of the century, electric lamps were being made for many specialized applications, including dental and medical. Safety lamps had been devised for use in mines and early advertisers were using filament lamps in their new electric signs. Special street lighting or municipal lamps were developed to compete with arcs, but incandescent lamps on series circuits were not truly satisfactory until produced with tungsten filaments.

In 1892 incandescent and arc lamps were installed temporarily to illuminate the Statue of Liberty during the Columbia Exposition. To the public, the most spectacular feature of the Exposition itself was the marvelous outdoor illumination of the buildings, electric fountain, and grounds. It was "a veritable fairyland at night." Inside the buildings the electrical manufacturing companies had a chance to illustrate their lighting methods and equipment.

These are abstracts from a book by Philip Atkinson:[1]

> The arc lamp is adaptable for lighting large areas requiring powerful light . . . such as outdoors, workshops, stores, public buildings and factories, especially where colored fabrics are produced . . . in ordinary dwellings or for offices, a softer, steadier, and more economical light is required. Tests of four leading incandescent systems show an average electromotive force of 91.91 volts and a current of 0.71 ampere for 16-candlepower lamps . . . working at 4.07 watts per candlepower has been found to be the most efficient and economical. . . The present lamp charge . . . allowing a rebate for the platinum wire in an old lamp, is about 70¢ . . . Life can be added to a lamp . . . if the ends of a broken filament can be interlocked by jarring the bulb. . .

This reference included a note that the Paris Electric Conference in 1881 adopted, as a standard unit, the light radiated by a square centimeter of pure melted platinum at the temperature of congelation. Instructions for calibrating lamps or candles in terms of this reference standard were included and the Bunsen photometer was described.

SCIENTIFIC INSTITUTES

In the early fall of 1884, the Franklin Institute held the first all-electrical exhibition in the United States. It was in sharp contrast to the fiftieth anniversary celebration of the Institute just 10 years before. Then there were no dynamos, no electric lights, no suggestion of the revolution on view in 1884. Meantime, Thomson had made his reputation as a lecturer and was well received by such men as Sir William Thomson, who became Lord Kelvin in 1892, Lord Rayleigh and Professors Silvanus P. Thomson, and William Ayrdon. It was at this meeting that the ampere was finally established as an international unit, that a national bureau of physical standards in Washington was proposed—it took 26 years to persuade Congress of the need—and the ground work was laid to found the American Institute of Electrical Engineers.

INFLUENCE OF LIGHTING ON OTHER EQUIPMENT

The first dynamos were built to operate arc lights. As the sources were improved, the machines were modified to keep pace. Constant-current and, usually, d-c designs were employed. Arc lamps were best suited to series operation which led to the development of series-shunting and current-regulating devices.

Brush and Thomson were leaders in this field although Brush contributed more to the design of the arcs and Thomson specialized in the development of associated electric equipment. In 1879, to meet the competition of Brush's 4-light dynamo, Thomson designed a 3-coil, single-armature machine. He placed a 3-section commutator on one end and three collector rings on the other. The machine would produce direct current which the customer wanted and alternating current which appealed to Thomson as having the greater future. A variety of arc lights and distribution systems were successfully operated from the unit. The dual aspect of the machine established Thomson as a pioneer in the design and production of a-c equipment.

Edison reasoned that constant-voltage and multiple-distribution systems would be superior for incandescent lighting. Accordingly, he designed his lamps for high voltage and low current, and set about to design and build constant-voltage generators. In spite of expert opinion and proof as to the impossibility of paralleling such machines, he did so in 1880 when three dynamos were operated together on the steamship *Columbia*. A fourth machine was used as an exciter.

Circuits on this ship had fuses, but the only instrument was a pilot light at the dynamos. The operator gauged the voltage by lamp color. Fusing of electric circuits was conceived by Edison. He used lead wires which melted and interrupted the circuit in case of overloads or short circuits. These were used not only to fuse branch distribution circuits from the main but also on customers' premises. Many also contended that underground distribution was impracticable because of water leakage but, again undaunted, Edison successfully developed tubing, joints, junction boxes, and other equipment for this purpose.

Edison's invention of a 3-wire distribution system saved millions in the cost of central stations and spurred their development. Although his lamps were designed to operate at a relatively high voltage, the distribution of power at constant voltage resulted in high current in the distribution mains. The system he developed operated two 110-volt dynamos in series to produce 200 volts. Using a third or neutral wire between the two dynamos permitted the use of 110-volt lamps and a 60 per cent reduction in the amount of copper required.

In 1880 the idea of separate lamps and sockets was conceived. Edison's first lamps were fastened to a wooden stand but later the bulbs were finished with copper strips which made contact with similar strips in wooden sockets. The idea of the screw base was developed in 1880 and, although the length of the threaded portion was increased in 1881 and 1888, the design was later adopted by all American lamp manufacturers.

Edison's first meter (1882) indicated the amount of electric service rendered to a customer. Zinc electrodes in an electrolyte were fixed to a beam balance. The passage of current (a fraction of the total) caused the transfer of metal from one electrode to the other until the beam tilted. Connections to the electrodes then reversed and the switching operation was recorded. Service was reported in "cubic feet" indicating that the lighting supplied was equivalent to burning that quantity of illuminating gas. In 1887 Edison developed a voltage indicator for use in central stations although the pilot lamp was still retained. Shortly afterwards, he developed an ammeter which consisted of a coil of wire around a soft iron core. The distance the core was drawn into the coil indicated the current through the device.

Weston, too, designed generators and meters. His first generators were for electroplating, but the later ones were for arc and incandescent lighting. His greatest contributions, however, were in the field of instrumentation. In 1886, he designed a voltage indicator for lighting circuits which employed pole pieces cut in the form of a hollow cylinder. The current-carrying coil was placed on a pivoted soft-iron core and could move freely between the pole pieces. The degree of rotation indicated the current in the coil and the voltage being measured. At this time, Marcel Duprez and A. D'Arsonval in Paris were developing a galvonometer with the same form of field. Nearly identical discoveries and developments often were made simultaneously by independent investigators widely separated from each other.

Two years later an ampere-hour meter was developed by O. B. Shallenberger as the result of an accidental occurrence and the insight and ingenuity of this Naval Academy graduate. During the testing of an a-c arc lamp, a small spring dropped onto the plate of the actuating coil and began to rotate. Analyzing the cause of this rotation, Shallenberger was struck with its possibilities in meter design. A month later, he had constructed a successful ampere-hour meter.

About 1889 Thomson set out to design a mechanical meter with easily read dials that would record watt-hours or total electrical service. Thomson's meter was a tiny motor driving a Faraday eddy-current disk located between permanent magnets. The armature was connected across the line to respond to voltage and the field coils were energized with a portion of the current supplied to the customer. The rate of rotation was proportional to the load while dials that counted the motor's revolutions introduced the time factor for measuring total energy. Thomson's meter was completed in 1890, was awarded a 5,000-franc prize (Professor Arons received an equal amount for a meter which proved less practical), and was used as the criterion of excellence in judging electric instruments at the Columbian Exposition in 1893.

SWITCHBOARDS

To control lighting circuits, switchboards became necessary. The first ones were wood. Circuits terminated on wooden panels and even fuse links were exposed directly to their wooden bases. By 1889 few insurance companies would provide fire coverage for electric generating properties or large consumers. In this year, the Edison General Electric Company installed a board in Greenville, Pa. The power terminals were mounted on slate. Thomson-Houston shortly afterwards combined switches and instruments on slate and marble panels. It built the first completely fireproof board in 1892 for controlling the lighting in Carnegie Hall, New York City. Westinghouse built its first fireproof board for display at the Columbian Exposition. From it were controlled the 12 generators and 40 distribution circuits for lighting the fair grounds.

APPLICATION ENGINEERING

In 1894 F. W. Rane used glass-enclosed arc lamps several hours each night to provide additional light in a greenhouse. Some of the results on plant growth were desirable and some were not. This study originated the use of electric lighting for such applications but also launched many other investigations to begin the process of untangling and identifying the effects of radiation on living matter.

A year later, in 1895, N. R. Finsen started to study the action of radiant energy on the general health of the human being and its effectiveness in the treatment of disease. Skin infections were of particular concern and he used arc sources with quartz lenses to concentrate light and ultraviolet energy on local areas of the body.

Reflectors and lenses to control light were introduced much earlier than this period, but their use and variety had been quite limited. Spherical reflectors were first considered for light houses in 1780, while parabolic types were introduced about 1790. The latter were employed in a Parisian theatre as early as 1846. Simple convex lenses were known very early and in 1822 Fresnel introduced his thin lenses of stepped elements. Mangin developed a glass reflector in 1874 with spherical inner and outer surfaces of differing radii. The outer surface was silvered so that by reflection and refraction the effect of a parabolic reflector through a fairly large solid angle was obtained.

An English text of 1892 by S. R. Bottone,[2] contained these recommendations: "As a result of many and careful observations, . . . to light a room brilliantly there should be a 16-candlepower light to every 33 cubic yards . . . or a 20 candlepower lamp to every 40." A table for 10 candlepower lamps showed that the number varied with the proportions of the length, width, and height of the space. A 1,000 candlepower arc was considered sufficient for 750- 1,000 square yards in cotton mills, 1,500- 2,000 square yards in engineering shops and 20,000- 30,000 in street lighting. The analysis of lighting costs concluded with the summary: "We may say the cost of electric lighting at $7^{1}/_{4}$ pence per unit would be equivalent to gas at 6 shillings per 1,000 cubic feet or just double the average price of gas."

The proponents of electric lighting, particularly those using filament lamps, were quick to emphasize certain advantages for the electric. Gas lights were generally turned upwards while the electric could be hung downward, pendent fashion. Electrical designers claimed that this avoided all shadows and "obscured" lighting. There was some tendency to identify single-lamp units as pendants and multilamp units as electroliers. The literature proudly announced that as many as 30 and 40 lamps were available on some electroliers. Many of the first ceiling units were on flexible cords with a pulley for height adjustment. Wall brackets with one or two lamps and some with flexible arms, in imitation of gas lighting, were also used. Desk and reading lamps, quite similar to many modern units, were available. Cut glass, silvered metal, and china accessories were featured for use in stores and other interiors.

Reflectors were first used for decoration but were soon designed to direct more light toward the work, onto displays, and at other objects or areas to be illuminated. Still later, the reflectors were deepened to shield the light sources and keep the light from going where it was not wanted.

Everly L. Haines conceived the idea that silvered-glass reflectors would be efficient, durable, and adaptable to the increasing variety of applications. His idea and reflectors appealed to Augustus D. Curtis and Arthur Morgan who reorganized Haines' small business into the National X-Ray Reflector Company. In 1897, the first mirrored-glass reflectors designed specifically for show window lighting were installed in the Washington Shirt Company haberdashery in Chicago.

Blondel, in France, applied the principles of reflection and refraction by means of glass prisms to the design of open reflectors and enclosed units. These, too, controlled the distribution of light from incandescent lamps and made electric lighting more effective. In 1898 the Holophane Company was organized to produce, market, and expand the applications for this type of equipment.

It was in the pioneering period of 1878–1902 that investigations into the proper application of light sources were begun on an organized, engineering basis. The transition from exclusive use of bare lamps to the adoption of light-controlling equipment gained momentum. Photometric laboratories were opened to develop new equipment and to accumulate empirical performance data upon which to establish general and specific design practice. The first laboratory was established at the Lynn Works of the General Electric Company in 1899. It was also employed for the routine testing of the luminaries manufactured there. In 1901, Ryan and Stickney introduced a practical photometer which facilitated the appraisal of lighting equipment performance.

Houston's and Kennelly's book[3] recognized the importance of surface reflection factors and pointed out their influence upon resultant illumination. It recommended 15-25 lux for reading printed pages and warned that less than 10 for such tasks would be fatiguing. Moonlight was reported as 1/8 lux and sunlight as up to 80,000 lux. Already nature's levels had been measured and were contrasted to indoor recommendations. (A flux is equivalent to about 1/10 footcandle.)

Comments by F. C. Allsop[4] indicated certain limitations in application practices that have been observed many times in later years:

> "As regards the selection and arrangements of fittings . . . no hard-and-fast rule can be laid down. . . . Getting the most effective lighting and pleasing appearance is no easy matter, and a qualification—judging by the eyesores to be met with every day—not possessed by every electric-light engineer. One comes across dining rooms, reception rooms, and so forth, with fittings more suitable for barn or work shop . . . shops with the lamps so badly arranged that it is impossible to properly inspect the goods owing to the glare . . . theatres, music halls, and other public buildings so inadequately lighted that quite a gloom is cast over the place. No doubt, in many instances, the proprietors of the buildings are themselves to blame, as many insist on having the lamps arranged to their own notions. . . . In the majority of cases it is only too evident that bad lighting is caused by carelessness or want of judgment on the part of the contractor. Seeing that nothing lends itself so readily to artistic adaptation as the electric light, this is much to be regretted."

THE LIGHTING PIONEERS CREATE AN INDUSTRY

In 1875, Charles Brush went to work for the Telegraph Supply Company in Cleveland. In the fall of 1876, this company began building his arc lights and dynamos as a side line. Within 3 years the firm was known as the Brush Electric Company. It's earliest installations have been noted. On June 30, 1879, the first central-station company in America came into being as the Brush-licensed California Electric Light Company in San Francisco. On March 21, 1880, the first municipally owned electric plant was placed in operation in Wabash, Ind. A Brush dynamo in the basement of the courthouse operated four arcs on poles above the dome of the building. The Brush Electric Light and Power Company was organized this same year to provide arc lighting on Broadway in New York City.

While a professor at Central High School, Elihu Thomson journeyed to Paris for the mechanical arts exhibition in the summer of 1878. He was intrigued by Jablochkoff's electric candles and a-c equipment but was also impressed by the quiet, steady operation of d-c arcs in a small exhibit in the Gare St. Lazare. He made the decision there that his future was to improve "lighting by electricity."

Upon his return to the United States he told Edwin J. Houston, his colleague on the high school faculty, of the decision. The two had made an agreement some years before to work together and pool their developments, patents, and royalties. During the winter, Thomson secured financial backing from Philadelphia friends and set about to design his 3-coil dynamo which was also a 3-phase alternator. A day or two after the initial discussion, design drawings were completed. In steps of a "few days" each, castings were delivered; machine work was done; the coils were wound by Thomson and the machine was assembled on a planer bed of W. H. Harrison's machine shop. It was early March 1879, when a belt from the overhead shafting was slipped onto the dynamo pulley, and a borrowed arc burst forth with blue light.

"This is most embarrassing," said Harrison, as cobwebs, dirt, grease, cracked walls, piles of scrap, and other details of the little shop stood out sharp and clear in the flood of light. No one else was so concerned for each knew that he was participating in a historical event. All night Thomson observed and tested his machine. The next day he designed and started to build his own arc light. That fall, a small factory was opened and the orders began to come in, about the time Edison was keeping his 40-hour watch over the world's first practical incandescent lamp.

The Edison Lamp Company was formed to make lamps at Menlo Park in the fall of 1880. By 1881 demand had grown so that operations were moved to larger quarters in Harrison, N. J. These became the headquarters for the Edison Lamp Works of the General Electric Company.

Another organization known as Bergmann and Company was established in 1880 to produce the sockets, switches, and other accessories Edison required. It, too, had to be moved to larger quarters shortly afterwards. The Electric Tube Company was organized to make equipment for underground power distribution, which Edison inaugurated. Dynamos and heavy machinery were made by the Edison Machine Works. A factory on Goerck Street in New York was obtained for this activity in 1881. By 1886, more space was needed and the Works were moved to Schenectady, New York. Around it developed the world's largest electric manufacturing plant and the headquarters of the General Electric Company—the world's largest electrical organization.

Lighting created the electric power industry. Arc systems developed first but, by the application of his prodigious energy and ability, Edison quickly brought the incandescent lamp into full practicability. Its contribution to the growth of the industry soon exceeded that of the arc light. In 1881 a public service company was established with Edison equipment by Appleton, Wis. The dynamos were driven by water power, making this the first "hydro" station. The New York Edison Electric Illuminating Company had been incorporated in 1880, but it did not begin operation until September 4, 1882. On that date, the famous Pearl Street Station began to serve 59 customers with electricity for lighting.

Edison's first dynamos were designed to operate at 1,000 rpm while steam engines ran at only 100. In planning the Pearl Street Station, he chose to drive 1,200-lamp "Jumbo" dynamos directly and was finally able to obtain 350-rpm 120-horsepower steam engines. Because of high land cost, boilers were placed on the ground floor of the building while the engines and dynamos were put on the second. A third floor was added to provide test racks for 1,000 lamps.

Quality control of product was of early concern to Edison, but lamp demand progressed much faster than manufacturing techniques. The origin of the electrical industry was so closely associated with lighting that the industry has never fully outgrown the impact of this situation in the early days. As has been explained, lamp resistance varied considerably. The resistance of finished lamps was measured, however, and lamps were grouped according to the voltage to produce rated candlepower. Central stations were set up at corresponding voltages to absorb all the lamps produced. A lower-cost product resulted but the utility companies became established with a range of operating voltages. As lamp manufacturing techniques became more exact, lamps were made uniformly for each of these voltages. Complete voltage standardization has not yet been achieved although the trend has been toward uniformity at 120 volts.

Direct Current Versus Alternating Current

In addition to the contest between arc- and incandescent-lamp proponents, there arose one between those who supported d-c systems and those who favored the a-c type. The Edison group had originated and developed incandescent lighting with d-c systems. George Westinghouse chose to back a-c developments in entering the electrical business. While most of the pioneers took positive sides, a few, like Thomson, designed equipment for both systems.

In 1881 two French inventors, L. Jaulard and J. D. Gibbs, applied for the first of their patents on a-c devices and systems. In 1884 they furnished arc and incandescent lighting through transformers in a number of stations on London's Metropolitan Railway and designed a transmission line as part of the Turin Exposition that year.

Westinghouse purchased American rights to the Jaulard and Gibbs patents in 1885 and established a laboratory for William Stanley in the latter's home town of Great Barrington, Mass. Stanley was one of the pioneers in incandescent-lamp and d-c investigations who became interested in a-c studies and designs. During the winter of 1885–86, Stanley installed and operated an experimental a-c system which served about 150 incandescent lamps in the stores and on the streets of Great Barrington. Further experiments were carried out in Lawrenceville, a suburb of Pittsburgh, Pa., and the first commercial installation was put into operation at Buffalo, N. Y., in November 1886. In 1889, the Willamette Falls Electric Company installed a high-voltage a-c system to supply street lighting in Portland, Oreg.

Just before Thomson had been plunged precipitously into manufacturing in 1879, he had designed a-c generators, transformers, and circuits. In making transformers, he had closed the magnetic iron path of induction coils to produce the first shell-type core. It was also an idea of his to place transformers across the lines of a-c circuits as well as in them, series fashion. His objective was, like Edison's, to improve electric lighting by the sub-division of large sources on series circuits into smaller, independently controllable ones in parallel.

Late in 1886, Thomson had the opportunity of resuming his studies of a-c systems. A small alternator and transformers were built to light a new factory building with incandescent lamps. They worked perfectly but there was more to be done. Some were condemning alternating current as being lethal while others were using it without appreciable concern for the consumer's safety. The low-voltage aspect of d-c incandescent lighting service was emphasized while the high voltage of series-arc circuits was ignored. Thomson believed the real opportunity for alternating current lay in even higher distribution voltages, but it must be safe and he made it so. Normally, transformers would protect consumers from these high voltages but not if transformer insulation broke down. The solution lay in grounding the secondary coil.

Thomson-Houston then entered the a-c equipment field. Converting an installation in Montpelier, Vt., to alternating current in 1887, enough copper wire was removed from the original feeders to cover the cost of doing the job. A-c systems were shown to be safe, economical, and practical. In 1900, Thomson developed a constant-current transformer which was superior for operating arc lights on a-c systems than the earlier reactance regulators or pile drivers. Arcs had set the design of early generators as constant-current, d-c devices; this transformer made their use on the better, constant voltage, alternating-current machines more practicable.

Corporate Consolidation

In these years, nearly identical contributions often were made by more than one of the pioneers. In other instances, patent rights held by individual companies could not be fully utilized because of patents held by competitors. Litigation and infringement accusations were frequent. Most of the small organizations disappeared through consolidations or in bankruptcy during the panic of 1893.

The World's Fair of 1893 highlighted some of the competitive situations and patent difficulties which the industry faced. In October 1892, the Court of Appeals in New York sustained Edison's patent for the hermetically sealed, incandescent-lamp bulb. Westinghouse announced that his company would still fulfill its contract for lamps at the Fair. All the resources of his organization were concentrated on making a workable lamp from the wooden-stopper lamp of Sawyer and Man (1878) and the glass-stopper designs of Farmer and Maxim. On December 3, 1892, the result was exhibited. The Westinghouse stopper lamp was not equal in life to the hermetically sealed type but it did protect the contract with the Fair and it was needed for only a few months until the Edison patent expired.

George Westinghouse formed the Westinghouse Electric and Manufacturing Company in 1886 to produce electric equipment. He started incandescent lamp manufacturing in 1883 and purchased the United States Electric Lighting Company in 1888 to obtain lamp patents and additional manufacturing facilities. This company was started in 1878 to produce lamps under Maxim and Farmer patents and had merged with the Weston Electric Light Company in 1882. Weston's operation had begun in 1877 with arc lights and had added filament types in 1881. In 1888, Westinghouse purchased controlling stock in the Consolidated Electric Light Company and obtained the Sawyer-Man patents.

Along with his first orders, Thomson received an offer of greater financial support in the fall of 1879 if his plant was moved to New Britain, Conn. With the urging of his Philadelphia backers, the move was made and the American Electric Company of New Britain opened on March 29, 1880—Thomson's 27th birthday. Two years later, however, the company was "failing to develop, manufacture and promote its products with all reasonable diligence" and, by the provisions of their contract, Thomson and Houston turned in their stock shares and demanded the return of their patents.

Thomson was asked to come to Lynn, Mass., to meet a group of men in this shoe-making center who were interested in starting an electrical business. He reviewed his experiences, stated his faith in the future of the electrical business, and immediately won the support of the group which included Charles A. Coffin. An agreement was drawn and signed that day which resulted in the formation of the Thomson-Houston Electric Company in 1883. It was Coffin's suggestion that the firm's name should honor the man on whose work the success of the company depended. True to his personal agreement of years before, Thomson insisted that E. J. Houston's name be included.

The company advanced rapidly under the guidance of Coffin and with the equipment designed by Thomson. Both men surrounded themselves with other brilliant men and the company purchased many smaller manufacturing properties. In 1888 the Van Depoele Electric Manufacturing Company was purchased. It had been established in 1880 to produce arc-lighting and electric-traction equipment. Reorganized in 1882, it was continuously in financial difficulties. The Fort Wayne Electric Light Company was purchased in 1888 and the Schuyler Electric Company in 1889. Both made arc lights but were having financial troubles.

In 1884 Thomson-Houston bought control of the Consolidated Electric Light Company. It had been formed in 1882 from Sawyer's Eastern Electric Manufacturing Company which held rights to the Sawyer-Man incandescent lamp patents. The Sawyer-Man Electric Company which had been formed by Sawyer and Albon Man, his patent attorney, failed soon after its formation in 1878. In 1886, it was re-established by Consolidated and Thomson-Houston and the latter was licensed to make lamps under the patents held by the new organization. In 1887, Thomson-Houston sold its stock in Sawyer-Man to Consolidated, and in 1888, sold its interest in Consolidated to Westinghouse. It did retain licenses for manufacturing incandescent lamps.

In 1889 Thomson-Houston bought the Bentley-Knight Electric Railway Company and the Brush Electric Company. The next year, Excelsior Electric Company, another arc producer, was purchased. The growing company was capitalized at over $10,000,000 in 1891, rivaling the Edison General Electric Company.

Nearly each new phase of the growing electric light business brought another Edison company into existence. Whereas other pioneers limited themselves to one or, at most, a few of the expanding facets of the industry, Edison was in all of them. Companies for making electrical equipment, generating electric power, and promoting the use of electricity were the results of his boundless energy and genius to invent practical things. Many carried his name.

In a measure, Edison was too busy adding to his inventions and expanding his businesses to take time to contest apparent infringements of his patents. However, differences with Swan were resolved in 1882 and the English Ediswan Company was formed in 1883. That same year, the Edison Shafting and Tube companies were combined with the Edison Machine Works. In 1886, the parent Edison Electric Light Company absorbed the Company for Isolated Lighting and the Edison United Manufacturing Company was formed from the Lamp and Machine companies and the Sigmund Bergmann and Company.

The final step in the formation of the Edison General Electric Company came in 1889. It included the Electric Light and United Manufacturing Companies as well as the Canadian Edison Manufacturing, Sprague Electric Railway and Motor, and the Leonard and Izard companies. The new company was capitalized at over $12,000,000. The first president was Henry Villard who had headed the Oregon Railway and Navigation Company when Edison outfitted its steamship *Columbia*. In this year, suits were filed against all infringers of Edison patents and after 21 months of litigation, the courts handed down decisions that were a total victory for Edison.

Meantime, however, the company had sought funds from J. P. Morgan for expansion. Major Twombley investigated for the financier, and reported that he could recommend the loan only if the Thomson-Houston Company were absorbed by the Edison General Electric Company. Coffin was called to Morgan's office. Instead of being absorbed, the Thomson-Houston Company was merged with the Edison Company to form the General Electric Company in June 1892. The president of the new unit was Charles A. Coffin.

Coffin skillfully guided the new organization and the many Edison central station properties through the panic of 1893. Westinghouse, too, came through successfully but the next 10 years were difficult ones and only a few of the other electrical manufacturers survived until 1903. In this year Stanley, who had formed his own business after leaving Westinghouse, sold it to the General Electric Company and became the head of its Pittsfield Works. These consolidations had established organizations sufficiently broad to withstand the stress of a depression and protect the advances engineers had made in creating the lighting industry.

1903–1927 The Era of Research

The electrical industry expanded rapidly during these 25 years. The generation of electric power more than doubled four times. Over 100,000,000,000 kilowatt hours were produced in 1927 as compared with about 6,000,000,000 in 1902. Lighting progress was a lot steadier than in the first tumultuous years. There were still men of originality, ingenuity, and inventiveness. Through the co-operative effort of research, scientific and engineering knowledge was accumulated and translated into better products and services.

ARC AND DISCHARGE DEVELOPMENTS

C. A. B. Halvorson and W. L. Harraden, for example, developed Steinmetz's magnetite arc into a usable product in 1904. Some 300 units were installed immediately for street lighting in Jackson, Mich. Los Angeles, Calif., installed America's first ornamental street lighting system in 1905. Units containing seven lamps in circular globes were mounted "way up" at 13.5 feet. A. D. Jones in England combined enclosed- and flame-arc designs in 1908. At first cored carbons were used, but later solid flame carbons became available. The color of pure-carbon arcs was blue-white. In 1902, Blondel introduced salts to produce a yellow flame, and in 1910 Norman Macbeth produced white-flame photographic arcs. Just before the war, Beck in Germany developed high-current high-intensity arcs for searchlights. Cooling the carbons with alcohol concentrated the arc and intensified it. The original version of "white-way" street lighting was installed along Market Street, San Francisco, Calif. in 1916. Traditions in this city were such, however, that the effect was known there as the "Path of Gold."

Cooper-Hewitt mercury-discharge lamps were installed in the composing room of the *New York Evening Post* in 1903. This was the first commercial use of these sources but in the next 25 years they were employed in many industrial, photographic, special, and general lighting applications. Moore tubes were introduced commercially about 1904. Some 180 feet of tubing were installed in a Newark, N. J., hardware store and a second installation was made in the New York City Post Office. For many years Moore tubes were a lighting feature of the Engineering Societies Building in New York. Tubes filled with carbon dioxide produced a white light which was a close approximation to skylight. Although the efficiency was low, these sources were used for many years in color matching and grading applications. About 1911 neon tubing was introduced on a commercial scale and was used for decorative and advertising purposes. In the decade following 1920, low-wattage neon-glow lamps were introduced in Europe and in the United States.

INCANDESCENT LAMP PROGRESS

Incandescent lamps with osmium and tungsten filaments appeared in Europe in 1905. Osmium lamps were not produced in this country and the early use of tungsten was limited because a major improvement in carbon lamps was introduced at about the same time. The contribution came from Dr. W. R. Whitney and the research laboratory which he had established for the General Electric Company. This was the first such laboratory to be maintained by a commercial organization.

One of the early studies was into the cause of the bulb blackening of carbon lamps as they burned. Whitney believed that impurities might be the cause. If so, they could be removed in his new electric furnace which had been produced as a result of an earlier research project. The furnace produced a temperature of 3,500 degrees centigrade which was 500 degrees below the melting point of carbon and more than 1,600 degrees above the operating temperature of a lamp filament. Whitney had some filaments on hand. They were fired in the furnace and made into surprisingly good lamps. A repeat run with new filaments failed miserably. Finally the proper order of heat treating was worked out and an increase in lamp efficiency of about 25 per cent resulted. The filaments processed in the furnace were like metals in that they increased in resistance with increased temperature. The name GEM was used to identify them as General Electric metalized-filament lamps.

Dr. Werner von Bolton's invention in 1906 of the tantalum filament lamp resulted in the use of these sources here and abroad. Their efficiency was higher than the GEM lamps—about 50 per cent greater than regular carbon—but they did not come to be used extensively because of progress with other and better lamps. In the United States, their use reached a peak in 1910 when 3.5 per cent of all lamps sold were tantalum.

Tungsten filaments at this time were made by squirting powdered tungsten in an organic binder through dies. The binder was removed in an atmosphere of steam and hydrogen. Although the resulting filaments were extremely fragile, the superiority of tungsten as a filament material (7.85 lumens per watt for 100-watt 110-volt lamp) enabled them to compete with carbon. A 40-watt tungsten lamp was introduced in 1907 for $1.50 and a 60-watt size for $1.75. They sell for 14 cents—plus taxes—in 1952, a reduction of more than 90 per cent on a far better product. In 1908 the 25- and 100-watt lamps were announced, and in 1909 the 150-watt size was introduced at $2.50.

It was in 1908, that the Edison screw type of base was adopted as an industry standard. Originally, nearly every manufacturer used his own base and socket combination. By 1900, the Edison, Thomson-Houston and Westinghouse types predominated but confusion and inconvenience still resulted. Standardization was followed by increased mechanization of manufacturing, interchangeability and reduced cost. These, in turn, expanded the acceptance of incandescent lighting and increased sales.

In 1905, Just and Hanaman applied for American patents on their tungsten filament lamps. American rights were purchased by the General Electric Company. However, Dr. Hans Kuzel and Dr. von Bolton had also applied for similar patents and litigation again developed. The court's decision was in favor of Just and Hanaman who were granted patents in February, 1912.

DUCTILE TUNGSTEN FILAMENTS

When pure and cool, tungsten was very brittle; heating made it somewhat less so, but not enough to draw into wire for filaments. After more than 4, tedious, sometimes heartbreaking years of work in Whitney's laboratory, Dr. William D. Coolidge developed a process for making ductile tungsten wire. He had to originate methods of producing tungsten ingots, hammering or swaging them into rods while hot and, finally, drawing the rods through many stages of diamond dies. As he developed the process, he had to invent equipment to carry it out. It was another epic in persistence, patience, and ingenuity. In March 1910, he was able to announce success. A patent was granted him in December 1913, while, in 1914, he was awarded the Rumford Medal by the American Academy of Arts and Sciences.

Dr. Irving Langmuir joined the staff of General Electric's research laboratory in 1909. One of the problems being studied was the curious phenomenon of "offsetting" which caused short sections of filament to slip out of line with adjacent ones. In 1906, Lederer of Westinghouse's Vienna factory added thoria and other oxides to tungsten and found that offsetting was reduced. Langmuir undertook a study of the phenomenon to avoid introducing other materials into the pure tungsten filament.

GAS-FILLED LAMPS

Langmuir began by studying the gases driven from the filament while a lamp was burned. They were in such minute quantities that special apparatus had to be devised to make qualitative and quantitative analyses. The residual gases were found to be water vapor, oil (hydrocarbon) vapors, carbon monoxide, carbon dioxide, and hydrogen. When a small quantity of each was independently put into lamps, only water vapor caused the "bulb blackening" that was typical of regular lamps. Some of the other materials caused deposits, but only water caused the familiar black discoloration

Blackening was also shown to be the result of simple filament evaporation. Extremely high vacuum lamps were made and some were even operated with the bulbs immersed in liquid air to restrict contamination from the glass. They, too, blackened. The conclusion had to be that, except for water vapor, bulb blackening was caused by the filaments boiling away. This discoloration was the primary cause of light depreciation as a lamp was used so investigations were made into methods of retarding filament evaporation.

A tungsten-filament lamp was made with carefully dried and purified hydrogen at atmospheric pressure. It burned 360 hours without blackening but required considerable additional power to produce as much light as a comparable vacuum lamp. The gas was retarding evaporation but was also cooling the filament. Other gases were tried. Nitrogen was found to be inert toward the high temperature filament and conducted less heat away than hydrogen. At pressures above 5 microns and below 1, argon glowed as current passed from one filament leg to the other through the gas. It was found that this "Edison-effect," which caused the bulb to blacken rapidly, could be eliminated by using about 15-per-cent nitrogen with the argon. Heat loss from the filament was less than with pure nitrogen. Argon was available in only small quantities and further studies were carried on with nitrogen-filled lamps while research was started on ways to obtain argon from the atmosphere economically.

Langmuir made two sets of thick-filament lamps, one set operating in nitrogen at atmospheric pressure and the other set in vacuum. The filaments were operated at the same efficiency so that life results could be compared. The nitrogen-filled lamps were failures. It was expected that the circulating currents of gas within the bulb would cause some discoloration, but the deposit at the top of these bulbs was heavy and black, indicating water vapor. The experiment was repeated, with still greater precaution taken to eliminate all moisture. This time the gas-filled lamps were superior. The filaments, however, were 0.02–0.04 of an inch in diameter and required from 20 to 60 amperes. Such lamps, for 110-volt circuits, would consume more than 2,000 to 6,000 watts. By comparison, 25-, 40-, and 50-watt sizes were used in the home and the largest vacuum lamp then made consumed 500 watts.

Langmuir next conceived the idea that the effect of a large *filament* might be obtained by properly coiling a small one. The prior achievement of ductile tungsten wire by Coolidge made it possible. To obtain the effect of a heavy filament, it would be desirable to wind the coil on a large mandrel and have the coils close together. If too large a mandrel were used, however, the weight of the filament would cause it to sag and allow the turns of the coil to touch each other. Experiments showed what mandrel sizes and spacings gave the best results. Then Langmuir was able to make a gas-filled lamp of about 10 amperes (1,000 watts on 110-volt circuits) with the same life as vacuum lamps but twice the efficiency. Further experience made it possible to produce a 750-watt lamp which was put on the market in March 1914.

Langmuir had started out to study offsetting and ended up with the revolutionary development of the coiled-filament, gas-filled lamp. Meantime, an associate Dr. Aladar Pacz, had taken up the original problem. He believed that very pure tungsten would neither offset nor sag. Thoriated wire did not offset but tended to bend at high temperature. Pacz's work resulted in a filament made by an improvement in Dr. Coolidge's process that crystallized in a large over-lapping structure when first lighted. Both offsetting and sagging were eliminated and a better wire was available for the coiled-filament gas-filled lamps. The production of argon in quantity was also solved and argon-nitrogen-filled lamps from 200 through 1,000 watts were made available in 1914. By 1916, the 60-, 75-, and 100-watt sizes were listed.

SPECIAL SHAPES, FINISHES, AND DESIGNS

High-voltage lamps for circuits in the 220-volt range were introduced in the 100-watt size in 1910 while other wattages from 200 through 1,000 were added in 1915. Blue-tinted glass bulbs to produce "daylight" lamps were developed under the guidance of Matthew Luckiesh. The 100-watt and larger sizes were announced in 1917; 10-watt lamps in 1922 and the 25- and 60-watt sizes in 1926 completed this line. A different kind of quality was incorporated in the white-bowl type of 1920. Ward Harrison sponsored this feature to reduce the brightness of lamps exposed at the bottom of industrial reflectors. (A clear 150-watt lamp at that time cost $1.55; the white bowl sold for $1.70.)

The introduction of tipless lamps in 1919 improved the appearance and convenience of lamps and eliminated a cause of considerable breakage. Lamp exhaust and gas filling had been done through a tube which, when sealed off, left a small tip of glass on the surface of the bulb. Tipless lamps were sealed off at the other end where the tube could be covered by the base shell.

Inside frosting of lamp bulbs was perfected by Marvin Pipkin of the General Electric lamp development laboratory at Nela Park. Etching the glass was known to provide desirable light diffusion and brightness reduction, but bulbs so treated on the inside became fragile. Outside frosting had been employed but tended to collect dirt and grease and was less satisfactory in appearance and to the touch. A double-etching process on the inside of the bulb was finally developed and inside-frosted lamps were introduced in 1925. The bulbs were strong and smooth and, although intrinsic brightness was lower, there was practically no loss in light.

In 1912, advertisers were provided with standard-voltage metal-filament lamps designed for flashing service in electric signs. They were 10-watt lamps with clear bulbs. In another 12 years, 25-watt colored sign lamps were introduced. A 1,000-watt projection lamp opened the way in 1919 to more convenient projection equipment for advertising, entertainment, and instruction. A 500-watt lamp was introduced in 1925, this time with a "prefocused" base that put replacement as well as the first lamp in the correct position of the optical system.

Transportation systems with 32-volt power were provided with 15- and 25-watt lamps in 1909. Two years later 25-watt tubular show-case lamps were introduced. The 40-watt size was made in an even smaller bulb in 1914, the same year that the railroads and street railways began to use 100-watt lamps. Transportation companies also received a 100-watt lamp for interior lighting in 1917. This same year farm and country homes with their own 30-volt battery-generator power plants were lighted with 15-, 25-, 50- and 100-watt lamps. Through the efforts of electrical and mechanical engineers, home-sized generating plants brought light and power to many farms before wired service arrived. Some 2,000,000 "Delco" sets were made and, since they were often resold as utility service was provided, an even larger number of families enjoyed their use.

Electric lighting was replacing flame sources in other applications where central-station power was not available. Tungsten lamps for automobiles were introduced. The lamps on the Cadillac and Peerless models of 1912 were all electric—6 per car. Two automotive types introduced at 75 cents about this time have been reduced to 10 and 12 cents. (Taxes are additional. Until recently, automotive lamps were free of excise taxation while those for lighting homes, hospitals, offices, and factories were not.)

Tungsten-filament flashlight lamps were introduced in 1908. The first carbon flashlight lamps of 3.5-volt design were made about 1900—10 or 12 years after Gassner made the first practical dry cell. Manufacture of cells started about 1890 and had reached an annual volume of 34,000,000 by 1909. In 1908 all miniature lamps totalled 1,800,000, but this volume increased rapidly as the improved tungsten lamps came into use. By 1911 there were five sizes of carbon flashlight lamps ranging from 3.5 to 6.0 volts while tungsten lamps were made in four sizes from 1.0 to 5.0 volts. The cost of a representative lamp was 60 cents when introduced; it has been reduced to 12 cents. Radio-panel lamps for battery-operated sets were introduced about 1923. Since then, large as well as miniature filament lamps have been used by the millions for both functional and decorative purposes on radios.

During these years, filament lamps were made better, stronger, more efficient, and more convenient, and were reduced in cost as volume built up and engineers improved manufacturing techniques. The use of electric lighting expanded rapidly as more and more persons and more and more activities came to receive the benefits of electric light.

In reviewing the growth of lamp quality and value during these years, the introduction of a unique concept cannot be overlooked. In 1909 the General Electric Company adopted the trade mark MAZDA. It was later employed by the Westinghouse Electric and Manufacturing Company as a licensee. MAZDA was, however, more than a trade mark. Derived from the name of the Persian god of light, it was the symbol of advanced laboratory research and thorough product testing and inspection carried on by the companies and by the Electrical Testing Laboratories, independently.

SPECIAL APPLICATIONS

High-temperature and high-pressure quartz mercury-arc lamps were developed and used in England and on the Continent. As the years went by, they were used increasingly for research, water purification, and medical purposes. About 1913 a Miss Dudgeon in England used them to investigate the influence of light and radiation upon plant growth. They speeded germination and growth. Some plants bloomed and seeded early—desirable in some cases but not in others. The separate effects of infrared, light, and ultraviolet radiation had not yet been determined.

This was true, too, of the photochemical uses of radiant energy. Scheele had found, in 1777, that daylight reduced silver from certain silver salts. Wedgewood made black and white copies of paintings in 1802 but it took the work of Daguerre and Niepce to learn how to fix the images. This accomplished, the usefulness of lamps, particularly the low-pressure mercury sources with strong near-ultraviolet output, grew rapidly. Interest also developed in the use of infrared but more slowly and less widely. As early as 1914, engineers knew that filament lamps would cook food. As a stunt, they broiled steaks at a picnic at Nela Park in Cleveland that year. Although infrared was employed to some extent during the first World War, its true value and popularity were to develop later.

SPECTACULAR EFFECTS

The imagination of engineers in creating new sources was matched by that of those who applied them on increasing scales for display and decoration. For a period of 30 nights in 1907, Niagara Falls were floodlighted with arc searchlights. W. D'Arcy Ryan used arcs to floodlight the Singer Tower in New York in 1908.

With the introduction in 1914 of concentrated-filament gas-filled incandescent lamps, they were promptly applied to this type of design. The first large-scale installation was that on the Woolworth Building in New York. It was designed by H. H. Magdsick and inaugurated on New Year's Eve, 1914. Some 550 filament-lamp projectors brought out the architectural features of the tower and added glitter to the gold-leafed mansard roof. The lantern atop the building blazed with twenty-four 1,000-watt lamps behind crystal diffusing-glass panels. Automatic dimmers varied the brightness to gain attention and realism.

When France's freedom was threatened during World War I, popular interest was awakened in floodlighting the Statue of Liberty—a symbol of light and freedom to all the world and a gift of France to the United States. Previously, lamps had been placed within a row of windows cut into the flame of the torch and festoons of bare lamps had been placed about the statue during festivals. Magdsick was commissioned to design a more fitting installation. In 1916, 225 filament units were installed to floodlight the pedastal and statue. The original lines of the torch were restored as designed by Bartholdi. The flame was reconstructed of sheet copper with sections cut away to receive more than 600 pieces of yellow cathedral glass. The color was graded from dense at the bottom to light at the top. By means of optical elements and voltage control, the fifteen 500-watt lamps within the torch projected beams of light that gave a realistic flame effect.

Area floodlighting and relief lighting were shown on a large scale at the Panama-Pacific Exposition in 1915. This marked the first time that lighting engineers and architects worked together from the inception of building designs to the completion. W. D. Ryan's dignified but striking effects were in sharp contrast to the glare of festoons and outlines of bare lamps which had been used at previous exhibitions—and were retained in the Zone or amusement area of this one. A unique and dramatic feature of the Panama-Pacific Exposition was the "scintillator effect" produced by aiming the beams of massed searchlights through clouds of steam. A modern locomotive was used to produce the steam near the searchlight projectors.

In 1916, Chicago's municipal Christmas tree was decorated with "jewels" obtained from this Exposition. They sparkled in the light from banks of floodlighting projectors arranged around the tree. In 1917 J. A. Summers designed a flood-lighting system for the Capitol in Washington. The white dome dominated the restricted-height buildings surrounding it to be particularly effective.

THE FIRST WORLD WAR

By this time, both the Allies and the Germans were using light to assist their respective war efforts. Many new factors were introduced into this conflict; one was the airplane. A counter measure to daring night flying was the use of high-power searchlights. In 1918, American equipment was added to the French and English units already in use. The Beck high-intensity arc had been introduced in many countries. Sperry employed an air blast in place of Beck's alcohol jet to concentrate and intensify the arc crater. Beams of a 1,500,000 candlepower were produced with such arcs in units having parabolic reflectors of 5-foot diameter (focal length was about 40 per cent of the diameter to include 120–130 spherical degrees of source).

While searchlights were picking out night-flying aircraft, arc and filament lamps were serving in special services on war ships. Searchlights with narrow beams sent coded flashes from ship to ship by day and night. The transmission of 7 to 10 words per minute was possible with equipment having mechanical shutters in front of steadily burning sources. Work was done, too, on hydrogen-filled flashing signal lamps. The exceptionally high heat loss of this gas provided rapid darkening or nigrescence. Deep-red and ultraviolet signalling was also used to a limited extent on both land and water. Fluorescent screens were used to detect ultraviolet and ranges from 4 to 6 miles were obtained. Land communication was established over short distances by the use of portable sources which projected a pencil of light toward an observer who was equipped with binoculars.

In Italy some use was made of electric lamps submerged in the water to attract fish. This reportedly augmented the meager food supply and released more men for armed service. Commercial chicken hatcheries in this country made some use of artificial light to increase egg production.

When the war was over, celebrations and festivities were held to honor the returning soldiers and the dead. "Brighten Up" campaigns might well have been termed "Lighten Up." Lights blazed forth everywhere. Archs, memorials, and band stands glittered with festoons and were bathed in floodlighting. Searchlights pierced the night skies often augmented by steam for the scintillator effect. Main streets were relighted or adorned with decorative lighting effects.

Sight, Light and Vision

During the first years of the twentieth century, men began to study the process of vision and to determine the effects of external factors upon its operation. The first work of this sort was done by Koenig in 1897 who established fragmentary but fundamental relationships between visual acuity and illumination. As the years went by, more investigators—psychologists, ophthalmologists, physicists, and electrical engineers—contributed to the growing fund of knowledge. In 1906, the Illuminating Engineering Society was founded to give those interested in lighting an organization sympathetic to this branch of optical and electrical engineering and to "disseminate knowledge relating thereto."

For many years, the majority of the fundamental sight, light, and vision research done within the electrical industry was conducted at Nela Park in Cleveland, Ohio (the name was derived from the National Electric Lamp Association). The organization and physical plant were established in 1913. Located on the outskirts of a large city, it became a model for the research, development, and administration headquarters of many other industrial organizations.

Among those provided with extensive laboratory facilities was Dr. Percy Cobb. In 1915, he published results of studies on the relationship between visual acuity and pupil size and the influence of brightness and glare upon this relationship. Similar studies were carried on and published intermittently during the following years. By 1925 Cobb and Moss began to evidence their expanding interest in such matters by determining the influence various seeing and lighting conditions had upon the accuracy of performing certain visual and manual tasks. Their studies were conducted under various quantities of light and with light and dark surroundings which inaugurated "brightness engineering" research. Some years earlier, Dr. Matthew Luckiesh had become director of these activities and, under his leadership, the fundamental links in the chain of light, vision and seeing were being forged.

The demand for industrial production during the war years dramatically proved some of the early laboratory tests on the correlation among lighting, seeing, and productivity. New plants were lighted according to the codes of the day and, when old plants were relighted, records were often kept of the increased production rates. These gave emphasis to the need for better seeing through better lighting. It was in this period that artificial lighting became firmly established as a valuable tool of production.

Engineering Applications

Progress in the design of lighting equipment and the identification of application objectives are not as clearly and precisely traceable as lamp progress. There was important progress, however. In addition to the glass equipment already noted, porcelain-enamelled iron was used extensively for reflectors. The first industrial units were nearly flat and gave little shielding to the light source. Later, deeper units were produced and some were standardized in the familiar Reflector Lamp Manufacturers (RLM) shape. Others of the deep-bowl type provided even more shielding. For high-bay installations, better control of light than that afforded by porcelain enamel was required and prismatic-glass, silvered-glass, and polished or lightly etched metal units were developed. There also came into use a unit known as the glassteel diffuser that had a top-pierced porcelain reflector to allow upward light and an enclosing globe insert to reduce direct and reflected glare. The most typical practice during the war, however, was the use of 200-watt lamps in RLM-type reflectors on 10- by 10-foot centers. Codes called for 3–25 footcandles minimum in factories and 8–15 in offices. Preferred levels were 2–25 in factories with supplementary up to 100 on specific tasks. Good practice in offices called for 10–25 footcandles.

In 1908 Augustus Curtis introduced a revolutionary proposition of total source concealment or indirect lighting. He demonstrated the principle in this home before a joint meeting of the local sections of the Illuminating Engineering Society and the Ophthalmological Society. Silvered-glass reflectors were inverted to distribute the light over the ceiling, producing a low-brightness, luminous effect. *Electrical World* commented editorially: "The meeting at Mr. Curtis's home was perhaps the inception of the great movement of the present time for eye comfort and the conservation of vision by concealing light sources." Mr. J. L. Stair recorded his pertinent comment that "From that time forward light and illumination were no longer synonymous."

In 1920 an indirect adaptor for home-lighting lamps was made available. More and more designers specified indirect lighting for critical-seeing areas such as those in schools, offices, and drafting rooms. Indirect lighting was introduced into some store designs and was adapted in the form of coves and urns in the lighting of public buildings. Semi-indirect units were also developed. For example, enclosing globes of diffusing glass were modified to have dense bottoms and clear tops. Prismatic types—both open and enclosed—were designed for similar light distribution characteristics.

Lumen ratings were adopted for lamps to avoid the confusion that developed over candlepower values. Some manufacturers used horizontal, some spherical, while those producing large-area sources had to resort to equivalent candlepower ratings. Among the latter were Cooper-Hewitt and Moore who were the first to emphasize the quality aspects of low-brightness sources when used in other than indirect systems.

Blondel of France suggested the concept of luminous flux in 1894. It was defined first by the International Congress of Electricians at Geneva in 1896. In 1904 Hyde of the U.S. Bureau of Standards inaugurated a movement to define the "international candle" and to employ incandescent lamps in establishing and maintaining the value of the unit. With the support of the American Institute of Electrical Engineers, the American Gas Institute, and the Illuminating Engineering Society (IES) in this country, the proposal was adopted in the United States, England, and France. The value of the lumen was then established by simple geometric derivation.

About 1920 Ward Harrison and Earl A. Anderson announced the "lumen method" of illumination calculation. It was based upon the concept of experimentally determining the fraction of generated light which reached a 30-inch, horizontal workplane. They built a laboratory which permitted changes in room dimensions, reflectance values and types of lighting fixtures. After the laborious task of determining these fractions was done, the results were tabulated and published as coefficients of utilization. The data were further analyzed and empirical factors were developed for calculating the coefficients of other types of units.

There then developed the need for a convenient portable photometer for measuring illumination. There had been a steady development of laboratory instruments: the Ritchie in England, the Bunsen grease-spot unit and the Lummer-Brodhun prismatic type in Germany—from which the American Sharp-Millar and Macbeth instruments were developed. The last had been made flexible, convenient, and adaptable to a variety of photometric measurements but was still a precision instrument for experts. Then Dr. Clayton Sharp produced a simple device based on the slot or limit-gauge photometer by Trotter in England. It was further developed by C. F. Sackett into a compact "footcandle meter" of low cost, providing the first instrument for popular use in the measurement of illumination in actual installations. Batteries and flashlight lamps illuminated a series of grease spots of varying density which were placed along a scale calibrated in footcandles. When room illumination matched or caused a spot to disappear, the value opposite indicated the footcandles at the point of measurement.

There was no instrument or simple method of evaluating the quality of lighting installations, however. This quotation from an IES code[6] summarizes the problem: "There are so many factors entering into the situation (of glare) that it has not been found practicable to develop any instrument which will properly evaluate them all. To arrive at an intelligent judgement . . . the inspector must be reasonably familiar with the principal factors in or causes of glare."

Primarily to acquaint its own staff with some of these factors, the general Electric Company established full-scale lighting demonstration rooms at Nela Park in 1919. In 1921 the facilities were expanded and members of electric utilities and other elements of the industry attended the demonstration conferences in large numbers. The institute method of lighting education was born. Late in 1923 the operation was again expanded into the Nela School of Lighting to encompass the fields of lighting. Assistance was given in the planning of similar activities opened in London (1923), Paris (1923), and Berlin (Licht Haus, 1925) and other foreign centers. In this country, lighting institutes were opened in Boston, Mass., Harrison, N. J., New York, N.Y., Chicago, Ill., and Kansas City, Mo.

In 1927 this company also pioneered in establishing a field engineering service. This placed trained and qualified engineers in the major centers throughout the country to assist in the solution of lighting problems and to introduce the developments of the lighting laboratories into the commercial, industrial, educational, cultural, residential, and recreational activities of the land. As this period closed, the lighting industry and illuminating engineering were seemingly mature, stable, and advanced activities.

1928–1952 The Better-Light Better-Sight Era

The last 25 years of electric-lighting development defy a single descriptive interpretation, yet the period is aptly identified as one of better light and better sight. Sources became more efficient, more convenient, and lower in cost. As lighting and other uses for electricity increased and as engineering improvements were made in the design and operation of central stations, the cost of electricity was reduced. The public consumed the products and services of the lighting industry at ever increasing rates, during even the years of depression.

THE SCIENCE OF SEEING

About 1930 Dr. Matthew Luckiesh and Frank K. Moss consolidated the data from 20 years of work in the Lighting Research Laboratory at Nela Park. A wealth of information had been accumulated on light and sight and how the demands of the visual process affect the total human being. Research had explored, first, the influence of external conditions upon vision and, later, the expenditure of human resources in seeing. The co-ordinated data established the basis for a science of seeing. In the years that followed, the new science grew in content and enlarged the scale and perspective of lighting concepts in quantity of light and quality of lighting. The effects of ultraviolet and infrared on animate and inanimate matter and processes were studied, too, and application principles were formulated.

MOMENTOUS YEAR OF 1933

The year 1933 was a momentous one. The world was in a severe depression. The government in Washington inaugurated the New Deal. Commerce and industry sponsored the Century of Progress in Chicago and those interested in lighting launched the better-light better-sight program.

The exposition in Chicago dramatically demonstrated progress in lighting. Luminous elements of large and small scale dominated buildings and displays, pointing the way to closer integration of light in architectural design. Applied floodlighting was used extensively while the composition of many buildings emphasized the trend that had been developing in Europe and America toward the design of self-luminous structure. In 1930 the utilization of the technique had been advanced by Magdsick and Potter who introduced quantitative performance factors and systematized design procedures.

During a meeting of an Edison Electric Institute committee in the spring of 1933, George Whitwell of the Philadelphia Electric Company proposed that the lighting industry carry to the world the significant conclusions of the science of seeing. In so doing, it would render a valuable service to the public and would be contributing to business recovery. H. Freeman Barnes, an advertising executive with engineering training, was one of the leaders in the development and promotion of the better-light better-sight program. He and his associates presented articles, sound movies, and literally hundreds of talks to the industry and to the public in cities from coast to coast. Barnes was exceptionally adept and effective; his work is cited to represent the ever-expanding opportunities and contributions of those trained in engineering.

One of the most potent factors in the growth of the lighting industry and its maintenance of progress at this time was the institute method of lighting demonstration and training. In 1933, the Nela Park Institute was expanded and assigned an entire building. Visual interpretations of the better-light better-sight principles and program were made available for industry training. During the fall of 1934, five "travelling institutes" were sent out so that these demonstrations and conferences might be had first-hand in all parts of the country. Again engineers were instrumental in the preparation and presentation of these programs for training salesmen and saleswomen in better presentation methods and in better lighting designs.

Filament Lamp Developments

Other engineers were hard at the job of making better lamps. Old lamps were improved and brand new ones were introduced. In some cases, these changes came from new design and manufacturing techniques while, in others, they resulted from engineering studies into the needs of new or potential applications.

In 1934 the Corning Glass Works introduced a revolutionary ribbon-type bulb-making machine. Manufacturing processes had progressed from hand blowing of bulbs by expert glass-blowers (who later blew bulbs within standardized molds) to the Westlake machine that introduced mechanical blowing and, by 1926, produced 5,000 large bulbs an hour. The Corning machine delivered a continuous ribbon of molten glass from the furnace to a conveyor. As air jets from above caused blobs of glass to drop from the ribbon, they were quickly surrounded by molds in which the bulbs were blown to proper shape. The first machines produced about 15,000 bulbs per hour; later improvements increased the rate of production and the types and sizes of bulbs that could be produced. Marked cost and price reductions resulted as volume and quality of production increased.

Coiled-coil filament lamps were announced in 1936. For many years, single-helix filaments had been used and finally the difficulties of coiling the first helix into a second were solved. Further concentration of the filament in this manner minimized the cooling effect of the inert gas and increased lamp efficiency. This development in 1936 brought about a more complete realization of the objective that Langmuir had established in 1913 when he showed the advantage of using heavy filaments in gas-filled lamps.

Aviation matured in these years and lamps and lighting techniques were developed for its use. A 500-watt lamp for airway and airport lighting was designed in 1928. The 1,000-, 3,000- and 10,000-watt sizes followed in the next year. Other lamps were developed for use aboard the airplanes themselves.

In 1929 the lighting industry established by Edison's incandescent lamp was 50 years old. As part of Light's Golden Jubilee celebration, D. K. Wright developed a 50,000-watt incandescent lamp, the world's largest light source. To bring the 450-ampere leads inside the bulb, he modified the Housekeeper copper-to-glass seal which had been used in X-ray and other vacuum tubes. (Edison brought current into his lamp by imbedding platinum wires in the glass. The development in 1913 of dumet, a 2-metal wire of nickel iron sheathed in copper, improved Edison's method and reduced the cost.) Through Wright's contribution and the development of iron-nickel-cobalt alloys in Germany and here, high-current and high-wattage lamps became practicable. Heavy leads were brazed to metal ferrules which, in turn, were sealed to the glass. This construction also withstood appreciably higher temperatures than previous designs and, with the development of Pyrex-type of glass bulbs, led to the design of compact bipost lamps of high wattage-to-size ratios. Screens were developed, too, which collected the evaporated filament metal and reduced light depreciation due to bulb blackening.

In 1932 biplane-filament construction was introduced in projection lamps. Coil segments were aligned in two planes and so staggered that a more concentrated and a more nearly solid light source was presented to the optical system. Screen brightness increased 75 per cent. In 1933, a 100-watt projection lamp was made available; even lower-wattage lamps were introduced later. Had it not been for accumulated improvements in lamp and projector design, such sources would have been ineffective and of little use.

Three-light lamps having filaments of 100 and 200 watts which could be operated independently or in combination were introduced in 1935. Two-filament lamps had been made previously for a number of purposes, including that of flexibility in the quantity of light produced. These new lamps, however, were designed for use in home lighting portables as part of the continuing better-light better-sight program. Higher-wattage types of lamps were made for commercial installations.

Other changes occurred in lamps particularly useful in home lighting. Marvin Pipkin, whose inside-frosting process was applied to lamps in 1925, completed a continuing search for a better diffusion process in 1949. This one applied a silica coating to inside-frosted bulbs to further increase the diffusion of light, and again with no significant loss of light. The coating was used on a number of other lamps from 60 to 300 watts; it was applied uniformly on conventional, general-purpose bulbs and differentially on the reflector shapes. Another bulb shape was developed by Luckiesh and Commery to make lamps that would provide better lighting from the bare-lamp sockets and fixtures installed in millions of homes. A luminous, semi-indirect effect was obtained by enamelling the bottom of the bulb. It was designed with a concave contour above the maximum diameter to minimize bright reflections in the neck of the lamp and to minimize the area of the neck visible at usual viewing angles.

For many years, a better linear source had been needed for a variety of architectural and decorative purposes. The tubular showcase lamps were not particularly convenient and the color quality of mercury tubes made them inappropriate. The Frink Company had special designs made for its use under the name of Linolite lamps. In 1934 a simpler, less-costly construction was developed and Lumiline lamps were introduced. They were made in 30-, 40-, and 60-watt sizes with simple disk bases that permitted long, nearly uninterrupted lines of light. The first lamps were clear and showed how the filaments stretched the length of the tubes. Efficiency was low in comparison with conventional types, but the shape and introduction of colored, white-coated, and inside-frosted lamps soon led to their extensive use.

There were continued improvements in all lines of filament lamps. In 1933, for instance, prefocused automobile headlight lamps were introduced. They eliminated the need for focusing new lamps and maintained the lamps in proper reflector position in spite of road-shock and vibration. Convenience was added to the festival use of Christmas tree lighting by the introduction in 1934 of multiple-type lamps which overcame the disadvantage common to series circuits that all lights go out with the failure of one lamp. In 1935, toy-train lamps were produced; the demand for lamps on model equipment, toys, and entertainment devices later grew to millions each year.

The demand for flashlight lamps also increased rapidly. In 1934 some 60,000,000 were purchased, helping to create a market for 140,000,000 dry cells that year.

Some lamps went into the 200,000 electric trainmen lanterns in use, some into the 350,000 medical and dental appliances. In this same year, machine-made lens-end flashlight lamps were made available. They sold for 10 cents as compared to a price of $1.00 for similar but specially-made medical lamps. Prefocused flashlight lamps were introduced in 1937. These insured correct positioning of the initial as well as replacement lamps in focused-beam flashlight reflectors. Purchasers of previous lamps priced at 9–10 cents were quick to recognize the increased value of prefocused lamps at 13 cents. Few realized, however, to what extent the cost of power influenced the choice of lamp design factors, particularly the selection of life at a relatively few hours. Power costing 500–600 times that from electric utilities made high lamp efficiency essential.

In 1936 another improvement was made in radio-panel lamps. It was found that lamps were sometimes causing interference and further studies located the source of the difficulties in the mechanical crimping of the support wires over the filament. A welding process was developed and this particular interference of one electric device with another was eliminated. Lamps were used extensively on radios (some 7,500,000 in 1929 alone), but vibration caused many early failures. Filaments and supports often responded to different frequencies so that filaments broke loose or distorted and became ineffective. A line of lamps resistant to vibrations in the audio range was therefore developed. This was accomplished by positioning and proportioning the filaments and supports in such a way that they would move as a unit under vibration; a patent was filed by G. F. Prideaux in 1939.

For many applications, a major improvement in lamp value and usefulness was accomplished by applying reflecting metallic films to the lamp bulb. The early examples of internally and externally "processed lamps" were for limited specialty applications. Precise, high-volume manufacturing methods had not been sufficiently developed to take advantage of the light-controlling possibilities of such designs. In 1926, however, Silvray Lighting, Inc., inaugurated a lamp-silvering service with a process and materials that M. B. Beck had improved after purchasing rights to the basic technique.

The early workers in the field of inside silvered lamps included O. H. Biggs, Clarence Birdseye, and R. F. Strickland. Some time later inside silvering of discharge lamps was investigated and studied.

In 1936, silvered-bowl lamps were given national distribution in the United States. The processing charge, which had been $3.00 for a 300-watt lamp in 1928, dropped to 80 cents and, with still greater volume and improved processing, was reduced to 50 cents. Lamps in all the wattages commonly used for general lighting were gradually included in the silvered-bowl line. Neck-processed lamps with blown-bulbs of improved shape were introduced in this country during 1938. Designs have been expanded to include tiny, low-wattage lamps for lighting individual seats on airliners to large, high-wattage sources for high-bay industrial interiors. Some were developed with heat-resistant glass bulbs to withstand the high temperatures imposed by certain applications or the adverse weather conditions encountered outdoors.

The projector or sealed-beam-type lamp also was introduced in 1938. It represented an even greater achievement in light-control efficiency and precision. Being a rugged, heat-resistant lamp,. it also provided greater flexibility and convenience of use. The idea of this design originated with D. K. Wright in 1935. He inserted bipost base parts into a custard baking cup, aluminized the inner surface, mounted a filament within and sealed a glass cover to the cup. The "lamp" worked and research and development began on making reflector sections of high optical accuracy and covers with precisely designed and molded refractive elements.

By 1937, sufficient progress had been made that the idea of a large, all-glass sealed headlamp could be presented to the automobile manufacturers. In the next year, 150-watt projector lamps with spot and flood distributions were introduced for general lighting service and in August 1939, new cars with headlamps that did not grow dim rolled off the assembly lines. An idea born around a custard cup matured in the all-glass sealed-beam lamp. It offered such a forward step in precision of light control and increase of wattage with maintained performance that, for the first time, the entire automotive industry and state administrative bodies agreed upon a single standard of headlighting.

Processed lamps with aluminum and silver films for built-in control have been made available in many types. In 1940, sealed-beam lamps for trucks and busses were introduced. Beginning in 1941, they were developed for lighting and signalling in a variety of aviation and military applications. They were used in the slope-line aircraft landing system developed in 1949 to help keep aircraft operating in bad weather. Banks of lamps were so positioned as to show a smooth line of light to the incoming pilot as long as high altitude and alignment with the runway were correct. The pattern distorted if the approach was in error in either regard. Small lamps were developed for lighting the road ahead of bicycles, the platforms of railroad cars, and sculpture in museums. Large ones were designed for locomotives, displays in stores, floodlighting buildings, and operating photocells under bright daylight conditions to open plant gates or bus-terminal gates.

PHOTOFLASH LAMPS

Photoflash lamps introduced by General Electric in 1931 employed a new method of producing light. Produced under a patent issued to Johannes B. Ostermeier of Germany, they contained aluminum foil, an atmosphere of oxygen, and a short tungsten filament for triggering the flash. These enclosed lamps were far superior to flash powder in safety, timing, and light output and had advantages over continuous-burning sources for many applications. They were quickly adopted by professional and amateur photographers alike.

In 1935 engineers of the Philips Company in Europe introduced a photoflash lamp employing a filling of curled wire that replaced the previously used large-area pieces of foil of the Ostermeier-type lamp. The wire lamp produced flatter light output characteristics more suited to synchronization with certain types of camera shutters. Wabash Applicance Corporation obtained a license from Philips and put wire-type lamps on the market using a further development of the Philips Company relating to an aluminum magnesium alloy wire. Later, Rippl and Isaacs developed the shredded foil lamp which then replaced all other types of photoflash lamps with the exception of the SM or primer-types which were also a development of General Electric. In 1940 a blue-filter coating was applied to photoflash lamps so that they could be used with daylight-type color films. Light-absorbing filters were developed during World War II for infrared photography. Since then, other lamps have been developed with output and timing characteristics matched to new camera, film, and application requirements.

The introduction of photoflood lamps in 1932 resulted from a new application idea but required only an adaptation of well-established design and manufacturing principles. They were developed after Prideaux tried a 100-watt 64-volt train-lighting lamp on a regular lighting circuit. Lamp life was short (90 minutes) but the photographic effectiveness of a 1,000-watt lamp was obtained with only 250 watts. This line of lamps came to include several sizes with conventional inside frosted bulbs and a number of reflector types with spot, flood, and intermediate distributions. Efficiencies were adjusted to specific color temperatures for black-and-white and several types of color films.

INFRARED LAMPS

Similarly, the development of infrared and heat lamps followed a relatively simple course of lamp adaptation. Carbon-filament lamps were used from the very early days through the late 1930's. The Ford Motor Car Company was the first to apply large groups of such lamps to industrial drying and heating processes. Tungsten-filament lamps were introduced for such industrial applications after studies showed they had less depreciation, longer life and would be more uniform and better in over-all performance. Research had also established that short-wave infrared radiation penetrated skin and tissue more effectively than long. The higher color temperature of tungsten filament sources made them superior for numerous human comfort and animal husbandry applications. In 1951, Sylvania Electric Products Inc. announced hard-glass filament lamps for use as heating units in electric stoves.

DISCHARGE SOURCES

In 1932 Dr. H. E. Edgreton filed one of his early patent applications on a radically different photographic source. In principle it was not unlike Geissler's tubes, but in terms of design it illustrated the technological strides that had been made during the intervening years. An intense flash was produced by discharging large quantities of stored energy through the tube in a very short period of time. Other developments were contributed as the units came to be made commercially. During the war, a jumbo unit was designed to photograph many square miles of enemy territory from high-flying bombers at night. Postwar varieties included those small enough to surround a microscope barrel or to operate from portable power packs easily carried by a photographer. (Xenon has been used in most tubes). By varying the parameters of the energy-storing capacitors, the control circuit, and tube design, flash durations of a few microseconds and repetitive flashing for stroboscopic photography were achieved.

SODIUM-VAPOR LAMPS

In 1931 a glass was developed in Germany which would withstand the violent action of vaporized sodium, and in 1932 Dr. Giles Holst in Holland developed a low-voltage sodium-vapor lamp. Street-lighting installations were made soon afterwards in England and several European countries. Sodium lamps were introduced in this country in 1933 and the first installation was made on Balston Road in Schenectady, N. Y. The initial enthusiasm for this source in street lighting gradually diminished as the difficulties of designing effective lighting equipment for such a large-volume low-intensity source came to be recognized. The fact that sodium produced practically monochromatic light made it valuable for certain inspection and research applications.

MERCURY-VAPOR LAMPS

A trend toward higher pressures started with Cooper-Hewitt lamps before 1920 was resumed about 1930. Results awakened new interest in metallic-vapor sources here and, particularly, abroad. Lamps of nearly atmospheric rather than a few millimeters of pressure were described as high-pressure types, but the development of sources at many atmospheres soon necessitated a reassignment of relative values. On the new scale, the 400-watt lamps introduced during the early thirties by Philips in Holland, Osram in Germany, and General Electric Company, Ltd., were of moderate or intermediate pressure.

An 85-watt 3,000-lumen mercury lamp was developed here in 1933. Its first application was in street lighting on Post Office Square in Lynn, Mass. An auxiliary filament lamp was used to improve cold-weather starting and the color of light. In 1934 Sylvania used experimental 3-kilowatt high-pressure lamps for floodlighting its yards and buildings at Salem. Westinghouse was the first to announce a 400-watt lamp in 1934. The lamps introduced that year were constructed with an arc tube enclosed within a larger glass bulb. They were intended for general lighting and had a light output about equal to that of 750-watt filament lamps. Efficiencies were one-third less than sodium but the color was better and the smaller dimensions of the source could be accommodated to more efficient, effective, and economical luminaires.

A 250-watt mercury lamp was introduced in 1936 and was followed in 1939 by a more efficient one with inner and outer bulbs such as the 400-watt design. The latter has had extensive use in black-light applications. A smaller, higher-pressure 100-watt source was developed in 1938. It was made available within conventional bulbs for general applications, projector-type bulbs for spot and floodlighting, a filter-glass bulb for black-light uses, and special glass bulbs which transmitted sun-tanning vitamin-D-producing ultraviolet but absorbed wavelengths shorter than those found in sunlight at ground level. This sunlamp version was introduced by General Electric in 1939, just 10 years after the S-1 which was the first to employ such a carefully selected glass for the bulb that sunlamps could be used at home with reasonable safety. The S-1 employed a tungsten filament in parallel with a mercury arc to produce heat, light, and ultraviolet. In 1941 Westinghouse introduced a reflector-bulb sunlamp which contained a mercury-arc tube, starting switch, and a filament-type ballast. It was a "self-ballasting" lamp that could be placed in standard lamp sockets and, because of its simplicity and convenience, displaced nearly all other home-use types of sunlamps.

Wartime and post-war emphasis on industrial lighting expanded the mercury line of lamps. In 1942 the need for a high-wattage mercury source was met when a 3-kilowatt lamp was made available. Technological improvements lengthened the life, bettered the lumen maintenance, and increased the efficiency of the 250-watt and 400-watt lamps. A 1,000-watt lamp of similar design was also developed. The improved 400-watt arc-tube was employed in producing a color-modified lamp with a fluorescent phosphor coating within the outer bulb, a lamp with a reflector-type bulb and one combining the color-modification and reflector-bulb features.

COLD-CATHODE DISCHARGE SOURCES

Most of these mercury sources employed hot-cathode electrodes which were coated with materials that freely emitted electrons. As a result, these lamps could be designed for high-current at high efficiency. At the same time, cold-cathode sources were being used extensively as tubing in signs and decorative effects. Neon was used most frequently but in some instances other gases and mercury vapor were used to obtain other colors. Colored and even fluorescent glass tubing was used for the same purpose. A Frenchman, Jacques Risler, was one of the first to coat tubing with fluorescent powders but the phosphors were inefficient. Other cold-cathode developments were of limited but increasing utility in photochemical, photographic, therapeutic, sterilization, color-matching and research applications. Some of the more important examples were the Cooper-Hewitt lamps for photographic processes; the Hanovia and Uviarc mercury lamps for professional therapeutic purposes; and the sunlamps designed by Dr. Sperti. Low pressure mercury lamps for germicidal and bactericidal applications were worked upon by Clarence LeBel who filed patent applications in 1929. A number of investigators in this country, including Biggs and Burns at Sylvania, James and Rentschler of Westinghouse, did further work in connection with low-pressure mercury discharges incorporating activated cold cathodes for the production of 2,537 angstrom unit radiation employed in germicidal and bactericidal applications.

PRACTICAL FLUORESCENT LAMPS

Many persons contributed to the development of the practical form of fluorescent lamp. In 1852, Sir George Stokes discovered the basic principle of transforming ultraviolet radiation into light with materials like fluorite. Moving a quinine sulphate solution through the spectrum of sunlight produced by a quartz prism, he found it glowed when exposed to ultraviolet. Seven years later, Bequerel placed various powders inside Geissler tubes and found that some produced light. In 1896, Edison developed a fluorescent lamp and, shortly after the formation of the General Electric Research Laboratory in 1900, basic studies in discharge and phosphor phenomena were begun. By 1912 W. S. Andrews had constructed a working embodiment of a fluorescent lamp. In 1923 M. Risler gave a lecture before the French Society of Electricians regarding certain studies he had made in this field. In 1926 the Germans, Frederick Meyer, Hans Spanner, and Edmund Germer filed a patent application for an ultraviolet lamp containing a rare gas and mercury. The application also mentioned a Wehnelt cathode and added that if a coating of fluorescent material were used on the tube, "the lamp also could be used with advantage for advertising purposes." No attempt was made by the patentees to perfect the lamp.

In 1933 Dr. W. L. Enfield saw fluorescent-coated tubing used in Paris and received a report the following year from Dr. Arthur H. Compton of the progress being made in England with cold-cathode fluorescent tubing. Development work here was stepped up immediately to improve fluorescent sources that would be more efficient. Within 6 weeks, an 18-inch fluorescent lamp in a Lumiline-type envelope was produced. Phosphor excitation selectivity responsive to far-ultraviolet radiation of 2,537 angstroms produced by the low-pressure mercury discharge with high 2,537-angstrom generation was chosen; and the decision made to employ hot-cathode electrodes to:

1. Reduce cathode losses.
2. Reduce starting voltage.
3. Reduce operating voltage.
4. Reduce size and cost of ballasts.
5. Increase over-all efficiency.
6. Make short as well as long lamps practicable.

In addition, advances were made in standardization of dimensions and adaptability of high-volume mass production.

In 1935 experimental lamps were supplied to the United States Navy for shipboard tests. They were described at the IES National Conference in September 1935 and employed in lighting the banquet hall for a testimonial dinner meeting observing the centennial of the United States Patent System in 1936. Designs were completed in 1937 and the lamps were announced commercially before a joint meeting of the AIEE, IES, and New York Electrical Society in 1938. Early in the year, fluorescent lamps were installed in the Headquarters Building of the New York World's Fair. Even as the wonders of electric lighting were first shown at the Columbian Exposition, many of the public saw their first fluorescent lighting as it was used extensively at the New York and San Francisco Fairs in 1939.

The public received these new lamps enthusiastically. The desire for cool, efficient, practicable linear sources in color and controllable whiteness had been met. The first lamps were 15, 20, and 30 watts and designed to rational modular lengths of 18, 24, and 36 inches. During 1939 a 4-foot 40-watt lamp was introduced and large-area, general lighting was made practicable. At the same time, the white lamp was raised to a 3,500-degree-Kelvin color temperature; a 6,500-degree-Kelvin daylight color had been introduced initially. The RF lamp (rectified fluorescent of the Cooper-Hewitt type) was developed for industrial use.

The first fluorescent lamps employed a small choke for ballasting a single lamp. Cathode preheating was obtained with a manually operated switch. Later, automatic magnetic-vibrator and thermal-bimetal starting devices were developed. Of the bimetal types, the first were actuated by a heating coil and later by the heat of a glow discharge. The glow type was suggested by a student doing summer work with Westinghouse (and independently developed earlier in England.) Placed in parallel with the lamp, the low-wattage glow element became conducting. Heat from the discharge closed a bimetal switch to shunt the lamp and pass preheat current through the cathodes. On cooling, the switch interrupted an inductive circuit with a voltage surge sufficient to start an arc in the partially ionized lamp. Operating voltage of the lamp was low enough to prevent recycling of the starter. Lock-out devices which disconnected lamps and starters from the ballast at end of lamp life were later added.

In the fall of 1938, G.E. Inman, who had directed much of the development work on fluorescent lamps, realized the importance of using two lamps on a single ballast to obtain higher power factor economically and to operate the lamps out of phase with each other for stroboscopic correction. His proposal was similar to split-phase designs developed in Canada by Osborne and in Holland by Uyterhoeven. In the spring of 1939, such ballasts were introduced and their use was actively promoted by the utility companies to avoid wasteful operation of lighting loads at low power factor. During 1939 Sylvania entered the fixture manufacturing field with a 2-lamp unit having power-factor and stroboscopic correction.

In 1940 a 100-watt lamp was announced. Not quite as efficient as the 40-watt size, it did give more light per foot and reduced the number and cost of associated lighting equipment. In addition to mercury vapor, argon was used in fluorescent lamps to assist starting. By 1948 experimentation had shown that, by the use of Krypton, the light output of the 100-watt lamp could be equalled at about 90 watts. Other characteristics of the lamps were such that they could be used in existing 100-watt lamp installations. In 1945 a 40-watt "low-brightness" lamp was introduced with the same physical dimensions as these more heavily loaded lamps.

In 1940, the introduction of soft-white lamps gave further evidence of the possibilities in creating a wide range of colors in fluorescent lamps. This shade of white departed considerably from any associated with the black-body radiators. It was a flattering color, not unlike, the effect of surprise pink filters used in theatrical and display work. A 4,500-degree-Kelvin color introduced in 1945 was designed for true color rendition consistent with considerations of efficiency and environmental influences. At about the same time, Sylvania introduced a warm white color. In the latter part of 1946, this company introduced a 40-watt fluorescent lamp having a small trace of zirconium oxide in the coating used on the coiled-coil cathodes. Increased life was obtained, and in 1949 life rating for these preheat lamps was increased to 7,500 hours.

Meantime, cold-cathode fluorescent lamps in semistandardized sizes were introduced. Because of higher cathode losses and higher ratio of starting to operating voltages, they were made in relatively long lengths. They were generally operated in series on high-voltage transformers and were restricted to relatively low current. Later, further standardization was achieved and multiple-lamp operation introduced. The cost of light with instant-starting cold-cathode lamps was about 25 per cent more than with 40-watt preheat lamps.

About 1942, ballasts and sockets for instant-start operation of 40-watt fluorescent lamps were introduced by the France Transformer Company and by Kulka Manufacturing Company, respectively. Some users adopted instant starting even though equipment and power costs were increased about 10 per cent and the life of the lamps (designed for preheating) was reduced.

In 1944 General Electric announced instant-starting lamps containing triple-coiled cathodes. The active material in previous electrode designs was held within coiled-coil cathodes. With the triple-coiled type, more material was applied and held more firmly under the stress of high-voltage instant-starting service. These cathodes also were employed in the hot-cathode "slimline" lamps introduced during 1944 in 6- and 8-foot lengths for general lighting uses. Larger diameter slimline lamps in lengths of 4, 6, and 8 feet were introduced in 1949; the longer ones soon matched the economy and convenience of the 40-watt preheat lamp which had become the dominant source in the fluorescent group. For the longer lamps, difference between starting and operating voltages was such that there was little advantage for preheat starting. With a satisfactory cathode for instant starting available, long lamps of the hot-cathode type became practicable. By 1952 life ratings of 7,500 hours were achieved for instant starting as well as preheat types of fluorescent lamps.

The introduction of series or sequence instant-start ballasts was a noteworthy advance. In lead-lag designs, maximum voltage was sufficient to start either lamp independently. For sequence-starting series-operating ballasts, it was only enough to start one lamp while operating another. Some early designs caused appreciable reductions in lamp life but, through industry co-operation, corrections were made and the benefits of lower voltage, lower loss, lighter weight, and less costly ballasts were retained.

At about the close of World War II, public health authorities announced that beryllium, traces of which were used in phosphors, was toxic to some people. General Electric had developed and obtained rights to foreign developments of beryllium-free phosphors and made them available to the industry. No lamps with beryllium phosphors were made after June 30, 1949, and most manufacturing was converted somewhat earlier.

By this time, the various manufacturers had a bewildering array of white fluorescent lamps on the market; there were more than a dozen in all. In November 1949, General Electric announced a program of rationalism and simplification with which the user had to make only two simple choices:

1. Did he wish a warm or cool atmosphere? The former traditionally provided by ordinary tungsten filament lamps; the latter comparable to daylight at noon.
2. Did he wish maximum light output with reasonable color rendition, or the best color rendition with reasonable efficiency of light production?

Four lamps were identified as meeting these inclusive and simple requirements:

1. Standard warm white—highest efficiency, reasonable color rendition.
2. De luxe warm white—fine color rendition, reasonable efficiency.
3. Standard cool white—highest efficiency, reasonable color rendition.
4. De luxe cool white—fine color rendition, reasonable efficiency.

The development of efficient red phosphors made possible the two de luxe lamps and the development of this simplified program of lamp color selection. It has rapidly increased the demand for lamps within this group.

Saturated colors remained practically unchanged and their use by architects, decorators and engineers increased gradually. Other phosphors were developed to produce black-light and sun-tanning radiations rather than light. In other instances, phosphors were omitted and a glass used that transmitted short-wave 2,537-angstrom radiation. These germicidal sources came into general use for air sanitation and industrial applications involving the protection of products from molds, bacteria, and similar contamination. Many other types and sizes of fluorescent lamps were produced for specialized, limited-demand applications.

In 1943 the lamp engineers who had so fittingly met the request for efficient linear sources with straight fluorescent lamps adapted them to a circular shape. A 32-watt 12-inch diameter lamp was made available in September 1945. Later, 8-inch and 16-inch diameter lamps were also produced and Westinghouse Electric Corporation introduced a semicircular type.

A further improvement in triple-coiled cathode design was employed in rapid-start lamps introduced in 1952. Continuously heated, low-voltage low-loss cathodes provided rapid starting without starters and more economical operation than instant-starting circuits. With voltages reduced, ballast losses were less and ballasts were smaller. The development was made available first in 40-watt and in 16-inch circline lamps.

The first fluorescent lamps had a life rating of 1,000 hours and efficiencies of 30–35 lumens per watt. As experience and testing were obtained, actual performance was found to be considerably better and it was improved continuously. By patient research and intelligent application of results, engineers raised the life, the efficiency, the lumen maintenance, and all other aspects of fluorescent lamp value. The rate of progress and the achievements accomplished surpassed those for any other source during any other period.

EARLY APPLICATIONS OF FLUORESCENT LAMPS

Use of these sources at the New York and San Francisco Fairs has been cited. In some early installations, they were used bare as novelties, as an improvement over bare incandescent lamps, or as power-saving equivalents of many enclosing globe units. There were some well-designed installations, too. As examples: the luminous ceiling and marquee for Zion's Co-operative Mercantile Institute in Salt Lake City; the Hotel Mansfield restaurant in New York with wall illumination supplied from elements recessed into the backs of the booth seats; Jaeckels Inc., fur salon with 450 thirty-watt lamps in shielded mirror lights, indirect units, and luminous ceiling elements; directional perimeter lighting in New York's Art of Tomorrow Gallery and International Silver Company display rooms, and in Pittsburgh's Whitney Museum of American Art; and the Lehigh Valley Railroad's dining car which was specially equipped with a 115-volt inverter for fluorescent lighting.

SUMMARY OF APPLICATION PRACTICES

The introduction of fluorescent lamps was bound to modify lighting practices significantly. Where filament lamps had been applied inefficiently for want of an extended light source, the fluorescent were immediately employed. For those designs requiring compact high brightness, the filament was retained and improved. In some areas of design either illuminant was suitable, but in most instances one or the other had predominant advantages. There were many applications in which they were combined to complement each other.

In summary, lighting applications followed practices established in previous years although certain trends were emphasized or more fully exploited through the development of new or improved sources and materials. There were also evident refinements in application techniques and clearer expressions of design and engineering concepts.

MATERIALS OF ILLUMINATION DESIGN

Tempering processes made glass less fragile and hazardous in case of breakage. Diffusion materials were improved and efficiencies raised. Photosensitive glass was developed and decorative and functional louver-type patterns were produced in flat and bent pieces. The arts and sciences of prismatic glass production were advanced and applied to lighting applications. Many new plastics were created; some were diffusing and some were crystal-clear, in which case they were used in prismatic-type equipment. Technological improvements gradually overcame many of the early difficulties which had been experienced with thermoplastic materials in fluorescent lighting fixtures.

Chemical engineering advances provided lacquers, enamels, and resin- or plastic-base finishes that improved lighting equipment and the utility of many lighting applications. Wire with thinner but better insulation also resulted and, with the approval of such material in 1941, more and better lighting was incorporated in existing buildings.

Aluminum was used in fixtures not only for weight reduction and ease of fabrication but for certain natural and developed properties of advantage in controlling light distribution and brightness. Electrolytic anodizing applied a hard, durable surface. An improvement in this respect and an appreciable increase in reflectance properties was accomplished by the Alzak process introduced in 1934 by Ralph B. Mason of the Aluminum Company of America.

INDUSTRIAL APPLICATIONS

Electrically plated-gold and anodized-aluminum reflectors were used during the thirties for radiant heating and drying tunnels in factories. Sealed units employing 500- or 1000-watt lamps reduced the glare of such installations while maintaining infrared efficiency. Later, processed lamps were used almost exclusively. A low-brightness general-lighting industrial unit known as the Silver Bowl Diffuser was developed with an Alzak insert above the lamp. Alzak reflectors were added to the prismatic- and mirrored-glass types employed in lighting high-bay factory areas. Neck-processed lamps also came to be employed in such applications.

By far the greatest use of mercury lamps has been in industrial areas. To some degree, they have been employed alone but most generally with filament for better color results and immediate resumption of lighting following a short-time power outage. Mercury lamps required time to cool and restart.

The primary value of the 40-watt fluorescent and 85-watt RF lamps was in general-lighting applications. Some 1,600 RF lamps were installed during 1939 in a windowless building built by the Austin Company for the Simonds Saw and Steel Company. It was one of the first large areas to be lighted with fluorescent sources. Another was a Cleveland lamp factory in which 40-watt fluorescent-lamp units were installed in continuous rows for the first time. The technique proved economical because the number of outlets was reduced and the fixture channels were used as wireways. Appearance was superior to that of many individual fixtures and the long rows improved diffusion and reduced shadows. Although the 1939 Recommended Practice of Industrial Lighting specified no more than 20 footcandles for general lighting, levels of 50 or more became common a year or so later.

About 1950 radical improvements in the quality of lighting employed in factories began to be recommended and applied. Fluorescent lamps had proved so much more comfortable and effective than filament that RLM fluorescent standards had been established with the same 13-degree shielding and closed-top design as filament units. But as time went on, it came to be recognized that many industrial areas warranted higher-quality lighting with more shielding and more upward light to soften contrasts and shadows. Discoveries that open units often depreciated less than closed ones helped gain acceptance for the trend.

INDIRECT LIGHTING

Indirect lighting became the predominant technique used in schools, offices, and drafting rooms during the thirties. As the demand for higher levels of illumination increased, the mark of a quality fixture was no longer its ornamentation but its efficiency and effect on visual comfort. Simple, functional units were designed for general-service and silvered-bowl types of lamps. Combination mercury and incandescent units were developed to take advantage of mercury's efficiency; many preferred the cool appearance of the mixture.

Fluorescent lamps were quickly employed for indirect lighting. Both opaque and luminous glass or plastic trough units were developed. George Ainsworth, a long-time proponent of indirect systems, fostered the plastic types while the Edwin F. Guth Company produced glass units which were used to relight the barracks at West Point. An indirectly lighted cove installed during 1940 in the Union Lodge Masonic Temple in Orange, N. J., demonstrated the control of light possibilities with fluorescent lamps. Alzak strip reflectors, designed by Curtis Lighting, Inc., projected the light of these sources uniformly across a large expanse of slightly arched ceiling.

COFFERS

Illumination levels gradually increased and, in time, design engineers became aware of complaints against the ceiling brightness associated with indirect systems. Relatively, the brightnesses were low and remained in the same ratios with other brightnesses at both high and low illumination levels. Observing that the area of ceiling brightness and the complaints were both reduced in rooms having exposed beams, engineers began to modify lighting equipment designs and certain concepts relating to the parameters of comfortable seeing. The introduction of coffers built into the ceiling was a result of this shift in design. The re-evaluation of principles led to placing more emphasis on the visual size or area associated with the highest brightness elements in the environment. Coffers reduced the area and order of maximum brightness in the visual field and improved direct-view seeing comfort.

TROFFERS

By chance, many of these coffers were about 4 feet square and 40-watt fluorescent lamps were used to replace filament sources. The linear form of the new lamps soon led to the adoption of trough-coffers or troffers as they came to be known.

Somewhat later, single-lamp aluminum troffers of semiparabolic cross section were introduced. These were derived from two noteworthy developments that overcame objectionable ceiling brightness with indirect lighting.

In the one instance, indirect lighting units were placed at the focal axis of etched-aluminum parabolic troughs several feet wide. Two such systems were installed at Nela Park in an office and in a conference room. In the latter, filament lamps were placed horizontally in a trough and covered by Fresnel-type lenses that produced a fan of light perpendicular to the axis of the trough and reflector. In this manner, reflector brightness was minimized for both crosswise and lengthwise viewing. In the second development, the principles of parabolic reflectors and Fresnel lenses were employed to produce the Parabolume Ceiling. Fitting flat against a structural ceiling, the device consisted of stepped elements of parabolic contour that directed light principally downward from an indirect unit placed at the foci common to the various sections. Aluminum troffers were evolved from these designs and the earlier white-finished trough-coffers and first installed commercially in the Chicago office of the A. C. Nielsen Company.

LUMINOUS CEILINGS

During the late twenties, luminous elements became popular and, following the Century of Progress, their use grew rapidly. Simplicity rather than detailed geometric ornamentation came to be emphasized in architecture and elements large enough to dominate the composition of an interior were employed. As wall-to-wall systems were developed, however, the brightness and conspicuousness of the elements diminished; direct and indirect lighting systems approached a common effect.

Glass was used in the early examples and was employed in the first fluorescent installations. One of these in Salt Lake City has been mentioned. Perhaps the first one to employ 40-watt lamps was installed in the First National Bank and Trust building in Grand Rapids, Mich. Somewhat later, corrugated acrylic plastic was used for the diffusing panels. The greatest interest, however, developed after studies conducted by Moon, Spencer and others at a number of New England educational institutions were published. Luminous ceilings were ideally suited to the experimental verification of illumination and brightness calculation methods these investigators had derived from theories of interreflections in infinite rooms. One of the materials employed in the studies was a corrugated, parchment-weight vinyl plastic which thereafter was used extensively along with acrylic plastics and improved diffusing and prismatic glasses in the design of luminous ceilings.

DOWNLIGHTING

Some of the earlier luminous elements and luminaires were equipped with reflectors, lens plates, and louvers to increase directional control of lighting and to reduce direct-view brightnesses. Anodized aluminum reflectors of spherical and parabolic shapes were introduced. Elliptical reflectors of this material were used in the development of improved theatrical equipment and recessed, high-wattage high-efficiency downlights. This reflector shape permitted sharp, patterned framing of beams from spotlights and downlights with small aperatures at the ceiling. Equipments of both types by Rambusch, Century, Kliegl, and others were used extensively in stores, theaters, churches, and public buildings. During these years, continuous-tap autotransformer types of dimming controls were introduced; they were employed for many lighting applications in these fields. Here, too, processed lamps came to be used in great numbers.

The cubistic decorative fashions of the twenties often added complicated metal grills and blades to luminous elements and downlights. In some instances, they effectively shielded areas of source and element brightness. Later, the functional aspects of the techniques were emphasized. Magdsick, Ketch, Potter, and LaWall investigated the influence of various design parameters on the efficiency, effectiveness, and appearance of luminous elements and downlights, and systematized design data for them. Following the introduction of fluorescent lighting, Potter, Darley, and Oetting extended the data to cover louvered designs which finally encompassed wall-to-wall louverall systems.

DIRECT-INDIRECT UNITS

In 1931 the Illuminating Engineering Society sponsored a program for improved study and reading lamps. The principle of direct and indirect lighting was incorporated by W. F. Little into the design of the IES Student Lamp. A number of manufacturers made these lamps which were purchased by the millions. The diffusing-bowl design became common in many table, floor, and pin-to-wall lamps. Later, studies by Commery resulted in major refinements in the shape of the bowls, the height of the lamps, and the use of shields to reduce the brightness visible over the top of the shades. Circline fluorescent and silica-coated incandescent lamps were also used to good advantage in the design of portable lamps.

For general lighting, incandescent direct and indirect systems were frequently combined, either with separate units for each distribution or with fixtures having upward and downward light. Of the units combining direct and indirect, most were enclosing globes of "general diffusing" classification. With the advent of fluorescent lighting, there was a far greater use of louvers and opaque side shielding, and direct-indirect units became the rule rather than the exception. These had restricted candlepower and brightness near the horizontal (the distribution curves were shaped like Gibson Girls) and were one of the most distinctive aspects of application progress resulting from the introduction of fluorescent lighting.

OUTDOOR LIGHTING

In 1941 a water highway 7.7 miles long was lighted; sodium units were placed 500 feet apart along both banks of the Cape Cod Canal in New England. Throughout the 25 years of this period, street lighting continued to receive extensive attention. Filament lighting systems were improved markedly through refinements in the design of refractive equipment, improved lamps, and the introduction of anodized aluminum reflectors and shields. Mercury came into considerable use as the years went by and, in England, a number of fluorescent installations were made. In America, extensive experimental work was done with fluorescent systems, but few commercial installations were made. Climatic and economic conditions and performance standards were different from those in England.

Different evaluation as to the importance of glare, silhouette seeing, and reflected-light visibility have led to similar situations as regards European and American automotive lighting practices. Comfort and glare elimination have been emphasized here. This led to the sealed-beam standardization program and heightened the interest shown in the polarized headlighting studies of Roper and Scott in 1941. Wattage requirements and accommodation problems during a transition period to polarized lighting seemed beyond practicable solution at that time.

There was a phenomenal growth in the use of floodlighting for sports activities after 1930. Within 10 years literally thousands of recreational, amateur and professional fields, courts, and pools were equipped for night-time use. Again processed aluminum reflectors came into use in addition to silvered- and prismatic-glass devices. Porcelain enamel was used in many low-cost units and processed lamps came to be used widely for the smaller installations. Many of these equipments and techniques were adapted during the war as production in shipyards and other outdoor areas went on around the clock.

LIGHTING IN WORLD WAR II

It is impossible to estimate accurately the degree to which lamps and lighting increased production, sharpened the effectiveness of our combat equipment, and shortened the war. As in civilian life, they were indispensable.

The 40-watt, 100-watt, and RF lamps were perfected when American industry girded itself for another wartime demonstration of its production prowess. Aircraft plants and other factories of previously unimagined area and height were effectively and efficiently lighted with these fluorescent sources. It some times took a bit of proving that diffuse sources could be effective when mounted more than 40 feet high, but in a typical plant about 4,000 feet long, nearly 200 units contributed to the illumination on top of large bombers. (Light from some 700 units reached spots on the floor before production started.) In contrast, tiny 5/8-inch diameter fluorescent lamps improved the inspection of shell cases, gasoline tanks, and other items of materiel. Other fluorescent, filament and mercury lighting systems played vital roles in production and administrative areas.

Lamps and application knowledge were channelled into many phases of the war effort. Seeing research led to a change in color used for low-level security lighting. Red lamps and filters replaced the traditional but more easily detected blue. Blinking glow lamps, operating at practically no drain from dry cells, marked safe paths through mine fields. Small lamps wired to batteries that became active in salt water saved the lives of many service men. Fluorescent and phosphorescent materials developed for black-lighted display and entertainment purposes were adapted to locate parachuted supplies, identify life rafts, and reveal the spot where an airman or sailor was afloat on the sea.

There were special lamps designed for illumination purposes. Dental and medical field units were issued lighting that equalled that in many a doctor's office or hospital. Lamps went aboard naval craft, airplanes, and land equipment in navigational operational, signalling, and recognition devices. Special signalling lamps went on and off quickly to give sharp, clear code flashes. Lamps with light-absorbing filters projected infrared beams toward the enemy. Reflected energy was picked up and turned into television-like pictures. Many of the enemy were killed at night because they were "seen" by our soldiers with such equipment. Actually, these devices gave our foot soldier as dramatic an advantage over the enemy at his range of combat as radar gave to the longer-distance operations of the Navy and Air Corps. Shock resistance to meet the conditions of combat had to be built into these lamps, too. Sturdy projector types of lamps were frequently used. Lamps and lighting played an important role in our successful conduct of armed conflict.

INSTRUMENTATION

New instruments were developed to increase the precision of engineering investigations and to establish practices in new fields on sound rational bases. Shortly after 1930, portable photocell illumination meters were introduced. German and American devices were developed at about the same time; the first here were produced by the Weston Electrical Instrument Corporation. In 1936 a reliable pocket-sized meter met the needs of those who did casual illumination survey work. The components were combined in an instrument smaller than a package of cigarettes in such a way that measurements of transmittance, reflectance, and brightness could be obtained conveniently. These and a 1932 version of the Weston instrument known as the Sight Meter were obtained by the thousands for both promotion and engineering purposes. The photocells frequently were employed in precision laboratory instruments which eliminated much of the time and error of visual photometry. About 1950 they were incorporated with recording meters and co-ordinated driving mechanisms in semiautomatic candlepower and illumination photometers.

Luckiesh and Moss announced a visibility meter in 1935. It was the first instrument that enabled the illuminating engineer, research scientist, and consumer to appraise the relative visibility of tasks. It also established a rational scale of illumination values in proportion to the difficulty of seeing tasks. Previously, the threshold of vision was the only condition available for quantitative appraisal; degrees of suprathreshold visibility were beyond measurement and specification. With A. H. Taylor, Luckiesh developed a brightness meter in 1937 and an electrostatic air sampler in 1945. The former was more convenient than former portable devices and covered a greater range of brightness values than most other instruments. Taylor developed a special instrument for measuring extremely low brightnesses in connection with wartime blackout studies. The air sampler was used for basic research and to determine the effectiveness of equipment and techniques in the field of air sanitation. Many former concepts were found to be erroneous as a result of more adequate measurements with this instrument.

SPECTACULAR INSTALLATIONS

Experimental installations providing radiation at high intensity or density have long been a mark of research and progress in this field. Often the exploitation of their promotional value has overshadowed their scientific value. For instance, an installation of 1,000 footcandles in the Lighting Institute at Nela Park in 1940 was, to many, just another stunt. It was the first step, however, in a series of developments which established electric solaria as an important adjunct to lighting progress. Later a 3,500-footcandle installation was made which produced the balanced radiations of heat, light, and ultraviolet in summer sunshine, even to absorbing specific infrared bands removed by the water vapor in the atmosphere. Less extensive installations have been made in the Senator Hotel in Atlantic City, N. J., to provide winter guests the same benefits of radiation those in the summer enjoy. Small units have been made available commercially for clubs, industrial plants, mine properties, and the like.

Another notable installation was that provided at the Beltsville station of the United States Department of Agriculture in 1951. Some 88 slimline lamps spaced just 1 inch apart provided a maximum of more than 2,000 footcandles. More secrets of why plants grow, flower, bear fruit, and seed were unravelled with this installation.

Electrical aspects of the Beltsville lighting were prophetic, too. A static, electromagnetic inverter supplied 360-cycle power to the lamps which operated at higher light output, higher efficiency, and without any stroboscopic effect or ballast noise. At this frequency, only small capacitors were required for lamp current regulation. By adjusting the saturating direct-current into the inverter, smooth control of light was obtained over a wide range.

Consolidation and Standardization

During this period, Sylvania Electric Products Inc. entered the fixture manufacturing business and greatly increased its volume of lamp sales. It was organized as the Hygrade-Sylvania Corporation in 1931 by consolidating the Hygrade Products Company (the present name was adopted in 1942). The roots of Hygrade Lamp Corporation extended back as far as 1901 to a business established for renewing burned-out carbon lamps. During the years from 1917 through 1930 Hygrade purchased many smaller companies.

In contrast to lamp manufacturing and, to some extent, utility expansion, there was little growth in the fixture manufacturing industry through consolidation. Many entered this field from others; by the thirties, many had gone out of business or turned to other lines. With the introduction of fluorescent lighting, the number of manufacturers again multiplied rapidly.

In 1945 license agreements between General Electric and other lamp manufacturers expired. For the many years that they were in effect, product standardization was reached within the industry before new lamps or modifications were announced to the public. Afterwards, orderly progress had to be sought exclusively through the channels of societies, industry organizations, and the American Standards Association. In most instances, however, consideration could be given to such matters only after products were announced for public consumption. For many factors such as fluorescent lamp colors, knowledge was insufficient for complete specification. This led to a condition of product variation slightly reminiscent of the industry's early days.

Standards of Lighting Quality

Standards of lighting quality were established in several codes and practices of lighting applications. Quantitative values of fixture brightness limitation were included in detail in a standard of classroom lighting[6] and in a broad generalization on recommended practice of office lighting.[7] Earlier, three other methods of comfort evaluation had been proposed as criteria for quality standards.

H. L. Logan introduced the proposal that the flux distribution under certain conditions of the outdoor environment be used as the criterion in rating indoor installations. He developed instruments and charting methods to record flux patterns in several arbitrarily but rationally defined zones of the visual field. Ward Harrison (later with O. P. Meaker) developed a glare-factor calculation system in 1945. It was based upon existing laboratory data and conclusions derived from experience where specific data were lacking. Subsequently, a special committee of the Illuminating Engineering Society analyzed existing data and established recommendations limiting the ratio of brightnesses within an environment to not more than 10 to 1 and preferably not more than 3-to-1.

Still later, the borderline between comfort and discomfort (BCD) concept was introduced by Luckiesh and S. K. Guth. Continuing research has more firmly established the relative influence on visual comfort of bright areas, the angular size of such areas, their position with regard to the axis of sight, and the general surround or adaptation brightness. These were the same parameters upon which the glare factors of Harrison and Meaker were based; in fact, their calculations have been adjusted as more precise and extensive data were developed. In the process of further research, Guth's work has taken on the aspects of a usable, practical system.

These studies have already brought more comfortable lighting to schools, offices, factories and other areas where human beings are employed at close, detailed, difficult seeing assignments for long periods of time. From them will come standards of practice that will obviate glare, the waste of human energy and, perhaps, the deterioration of sight now attendant upon the pursuit of these activities.

Horizons—Past, Present, and Future

Most of those who read H. G. Wells' *The First Man on the Moon* before 1938 were likely to have enjoyed, but doubted, his predictions of phosphorescent and fluorescent lighting. Some did not doubt; they already understood the basic principles and realized that the introduction of practical sources required only years of study, experimentation and patience.

Electroluminescence had been observed by Destriau in France, Andrews in this country, and by others as an obscure phenomenon whose very existence was sometimes questioned in technical publications subsequently. In August of 1950, electroluminescent lamps approaching practicality were demonstrated by Payne and Mager of Sylvania at the Pasadena Convention of the Illuminating Engineering Society. These lamps had initial brightnesses as high as 20 footlamberts at 3,000 cycles and about one-tenth that brightness at 60 cycles.

Engineers and public alike are aware that light is often produced by disturbances within the atomic structure of matter and yet unknown possibilities in the realm of atomic light sources will be explored. Little by little, more of nature's secrets of light production and utilization will be discovered. The light of the firefly, luminous fish, and types of vegetation may some day lead to a truly cold light that is commercially practicable. In the easing of seeing tasks, in the increase of vegetation, in the improvement of animal husbandry, in the perfection of inanimate processes, in the heightening of recreational enjoyment . . . in these and other ways, the application of light will contribute to the progress of mankind.

While the future holds these possibilities for all nations, light has come to be a servant only where electricity was abundant and inexpensive. Far more than half of the people in this world are still dependent upon flame sources. From them, the flickering, smoking, tenuous flames demand continual attention. In earlier times, light was so precious that this slavery to its production often became idolatrous worship. In our lives, however, electric lighting has become a servant and the nearly universal commodity.

In retrospect, the glory of the past century of engineering progress lay partly in the products and services developed but was even more in the attitudes of the men and women who made discoveries, created the products, and established the means of service. Long before this century, light became the symbol of enlightenment and freedom; in the history of our nation they have been defended vigorously. In looking into the future, the need is a reaffirmation of faith in these attributes. May those whose work is yet to be done find lighting an inspiration, second only to the preservation and strengthening of the freedom, initiative, enterprise, and integrity with which, under God, our predecessors labored.

References

1. The Elements of Electric Lighting (book), **Philip Atkinson.** D. Van Nostrand Company, Inc., New York, N. Y., 1889.

2. A Guide to Electric Lighting (book), **S. R. Bottone.** Whittaker and Company, London, England, 1892.

3. Electric Incandescent Lighting, **Edwin J. Houston, A. E. Kennelly.** *Electrical World and Engineer*, McGraw-Hill Book Company, Inc. (New York, N. Y.), 1902.

4. Practical Electric Light Fittings (book), **F. C. Allsop.** Whittaker and Company, London, England, 1892.

5. Code of Lighting Factories, Mills and Other Work Places. *Transactions*, Illuminating

Engineering Society (New York, N. Y.), volume X, number 8, November 20, 1915.

6. American Standard Practice for School Lighting. *Publication Number A23.1*, American Standards Association (New York, N. Y.), 1948.

7. Recommended Practice of Office Lighting. Illuminating Engineering Society (New York, N. Y.), July, 1947.

8. Fluorescent Lighting Manual (book), **Charles L. Amick.** McGraw-Hill Book Company, Inc., New York, N. Y., 1942.

9. La Lumiere (book), **A. E. Becquerel.** Didot, Paris, France, 1867.

10. The Electric-Lamp Industry (book), **Arthur A. Bright, Jr.** The MacMillan Company, New York, N. Y., 1949.

11. The Progress of the Fluorescent Lamp (book), **Lawrence Burns.** *Sylvania Technologist*, Sylvania Electric Products, Inc. (Salem, Mass.), volume III, January 1950.

12. Illuminating Engineering (book), **Francis E. Cady, Henry B. Dates.** John Wiley and Sons, Inc., New York, N. Y., 1925.

13. The History of the Incandescent Lamp (book), **J. W. Howell, Henry Schroeder.** Maqua Company, Schenectady, N. Y., 1927.

14. Menlo Park Reminiscences (book), **Francis Jehl.** Edison Institute, Dearborn, Mich., 1939. A detailed account of Edison's work by a close friend and co-worker.

15. *The Magazine of Light*, General Electric Company (Cleveland, Ohio), volume XVI, June 1947.

16. Artificial Light (book), **Matthew Luckiesh.** The Century Company, New York, N. Y., 1920.

17. Science of Seeing (book), **Matthew Luckiesh, Frank K. Moss.** D. Van Nostrand Company, Inc., New York, N. Y., 1937.

18. Seeing (book), **Matthew Luckiesh, Frank K. Moss.** Williams and Wilkins Company, Baltimore, Md., 1931.

19. The Rise of the Electrical Industry (book), **Malcum MacLaren.** Princeton University Press, Princeton, N. J., 1943.

20. Historical Sketch of Street Lighting, **Preston S. Millar.** *Transactions*, Illuminating Engineering Society (New York, N. Y.), volume XV, April 1920, pages 185–202.

21. The Lighting Book (book), **J. L. Stair.** Curtis Lighting, Inc., Chicago, Illinois, 1930.

22. Radiation, Light and Illumination (book), **Charles Proteus Steinmetz.** McGraw-Hill Book Company, Inc., New York, N. Y., 1910.

23. Artificial Light—Past and Present, **Walter Sturrock.** World Horizons, 1939.

24. *Transactions* and *Illuminating Engineering*, Illuminating Engineering Society (New York, N. Y.), volumes I–XLVI, 1906–51.

25. Beloved Scientist (book), **David O. Woodbury.** Whittlesey House, New York, N. Y., 1944.

No Discussion

Feedback Control Systems with Dead-Time Lag or Distributed Lag by Root-Locus Method

YAOHAN CHU
STUDENT MEMBER AIEE

A FEEDBACK control system is composed of a group of elements, which may be electrical, hydraulic, pneumatic, and so forth, whose behavior may be linear or nonlinear. The analysis and synthesis of a feedback control system with linear elements which can be described by a linear differential equation with constant coefficients have been well developed. In such analysis and synthesis, the term "transfer function"[1] frequently is employed to describe a servo element. It is a Laplace transform notation of the differential equation governing the behavior of the element.

Transfer functions are, in an analytic sense, the building blocks of a feedback control system. Quite a large number of linear elements may be described by the six basic forms[2,3] as shown in Table I. The first four basic forms are familiar. A few words on the last two forms may be worth while. The dead-time lag is a dead time during which there is no response. A useful explanation of the dead-time lag may be visualized by expressing it as

$$\epsilon^{-Ts} = \lim_{n \to \infty} \frac{1}{\left(1+\frac{T}{n}s\right)^n}$$

This means that a dead-time lag may be regarded as an infinite number of simple delays with the same time constant. In certain applications, a suitable choice of a finite value of n may give approximate representation. The distributed lag represents an element describable by a certain linear differential equation, such as heat conduction in a continuous medium (commonly used in process control), transmission of air through a long pneumatic line, concentration of a material in a diffusible solid, electric transmission line, and so forth. It has been shown[4,5] that the behavior of such elements is represented by the transfer function $\epsilon^{-\sqrt{Ts}}$.

Table I. Six Basic Forms of Transfer Functions

Name	Transfer Function
Gain or sensitivity	K (= a constant)
Integrator	$1/(Ts)$
Simple lag	$1/(1+Ts)$
Complex lag	$1/(1+2\zeta Ts+T^2s^2)$
Dead-time lag	ϵ^{-Ts}
Distributed lag	$\epsilon^{-\sqrt{Ts}}$

The notion of considering these transfer functions as basic forms is threefold. First, they are all describable by linear differential equations with constant coefficients. Second, they may be used to represent a large number of elements practically used by either using the basic form directly or by their combinations. Last, they can be analyzed and synthesized readily.

All the six basic forms given in Table I can be synthesized in the frequency domain. However, it is difficult to find a transient solution if the feedback control system involves the elements with dead-time or distributed lags. There are methods to find the transient solution, but they are either to approximate, or limited to very simple systems. This paper attempts to solve such problems by means of the root-locus method.[6] The original root-locus method first will be generalized by introducing the phase-angle loci and then be applied to the problems with dead-time or distributed lags. Examples will be included as illustrations.

Generalization of the Root Locus

The root-locus method was originated by Evans.[6] It allows design based on the transient response method. The essential point of this method is to plot the roots of the characteristic equation of a feedback control system on a complex plane with the gain of the system as a parameter. Thus, it gives a picture of these roots simultaneously with their corresponding values of gain. Transient solutions then

---

Paper 52-200, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952, and approved by the AIEE Committee on Technical Operations for representation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted March 11, 1952; made available for printing August 14, 1952.

Yaohan Chu is with Massachusetts Institute of Technology, Cambridge, Mass.

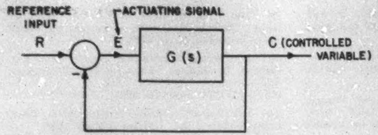
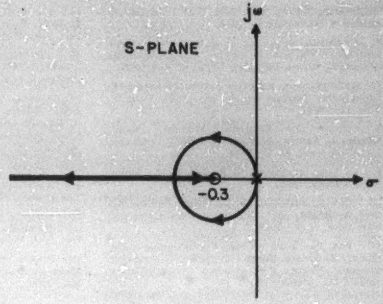


Figure 1 (left). Block diagram of a feedback control system

Figure 2 (right). The root locus for $G(s)=\frac{K(s+0.3)}{s^2}$

○ indicates a zero and × double pole



can be worked out once these roots are determined. In the following, this method will be stated briefly first, and then be generalized so that it may be used to solve systems with dead-time or distributed lags.

A feedback control system is shown in Figure 1, where $G(s)$ is the transfer function. It is well known that the characteristic equation of this system is

$$G(s)=-1 \tag{1}$$

This equation is complex and may be split into two equations by equating the magnitude and phase angle on both sides respectively. Thus, we have the magnitude equation

$$|G(s)|=1 \tag{2}$$

and the phase-angle equation

$$\text{Ang } G(s)=\pm 180 \text{ degrees} \quad \text{(for the principal plane)} \tag{3}$$

The plot of the phase-angle equation is called the root locus (of the characteristic equation), while the magnitude equation gives the value of the gain of each point of the locus. Figure 2 illustrates a root locus of a certain transfer function. Note that the locus is symmetrical with respect to the real axis. Equation 3 means that the root locus is a curve in which every point has a phase angle of ±180 degrees. The concept may be generalized by considering loci with other values of phase angle, or it may be written as

$$\text{Ang } G(s)=\phi$$

where ϕ is a constant phase angle. Thus, a family of loci can be plotted by various values of ϕ.

Figure 3 illustrates a family of root loci for the transfer function shown in

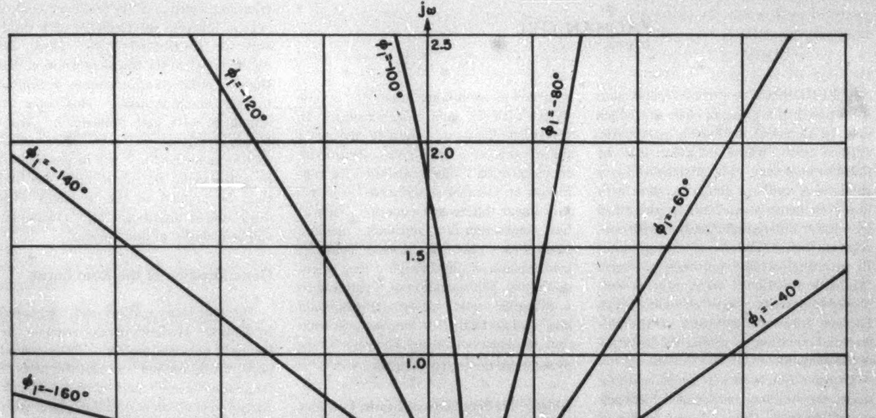


Figure 3. Phase-angle loci for $G(s)=\frac{K(s+0.3)}{s^2}$

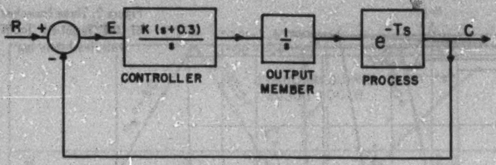


Figure 4 (above). Block diagram of a feedback control system with a dead-time lag process

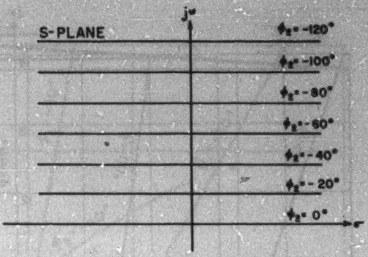


Figure 5 (right). Dead-time lag loci (T = 1)

Figure 2. Such a family of constant phase-angle loci for a given transfer function is called "the phase-angle loci." Since the locus with phase angle ϕ is symmetrical to that with phase angle $(-\phi)$ with respect to the real axis, only the loci on the upper half plane are shown.

Feedback Control System with Dead-Time Lag

If it is given a feedback control system with a transfer function $G(s)$ and with a dead-time lag ϵ^{-Ts}, where T is the dead-time, then its characteristic equation is

$$G(s)\epsilon^{-Ts}=-1 \tag{4}$$

The phase-angle equation is accordingly

$$\text{Ang } G(s)+\text{Ang } \epsilon^{-Ts}=\pm 180 \text{ degrees} \tag{5}$$

where $s(=\sigma+j\omega)$ is the complex variable. The phase-angle equation can be simplified into

$$\text{Ang } G(s)-\omega T=\pm 180 \text{ degrees} \tag{6}$$

A method is introduced now to obtain the required locus of equation 6. First it is to construct a family of root loci (or phase-angle loci) of the equation

$$\text{Ang } G(s)=\phi_1$$

Next, it is to construct another family of phase-angle loci (the name dead-time lag loci is suggested) on the s-plane from the relation

$$-\omega T=\phi_2$$

Then by superimposing these two families of loci on the same s-plane, the points of

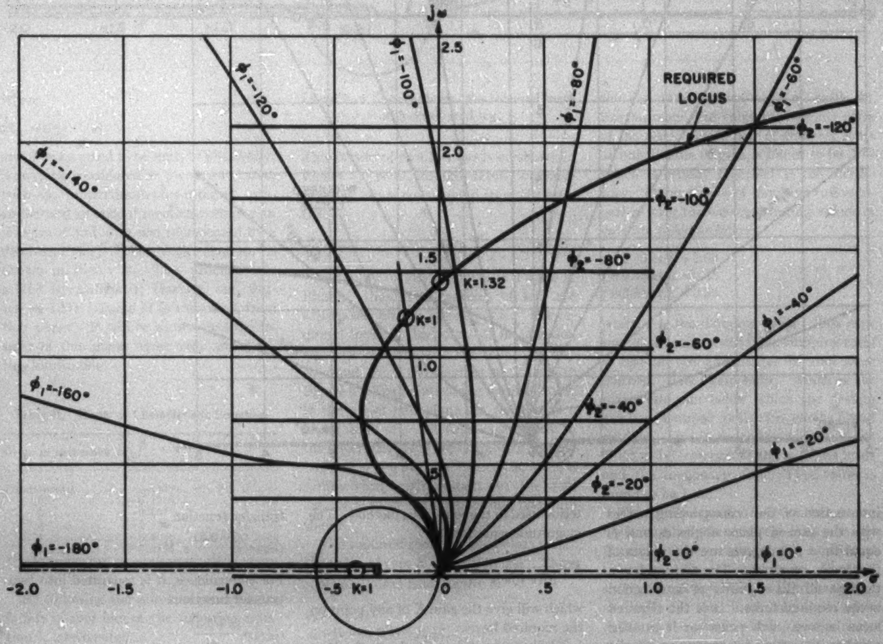


Figure 6. Construction of the required locus with dead-time lag

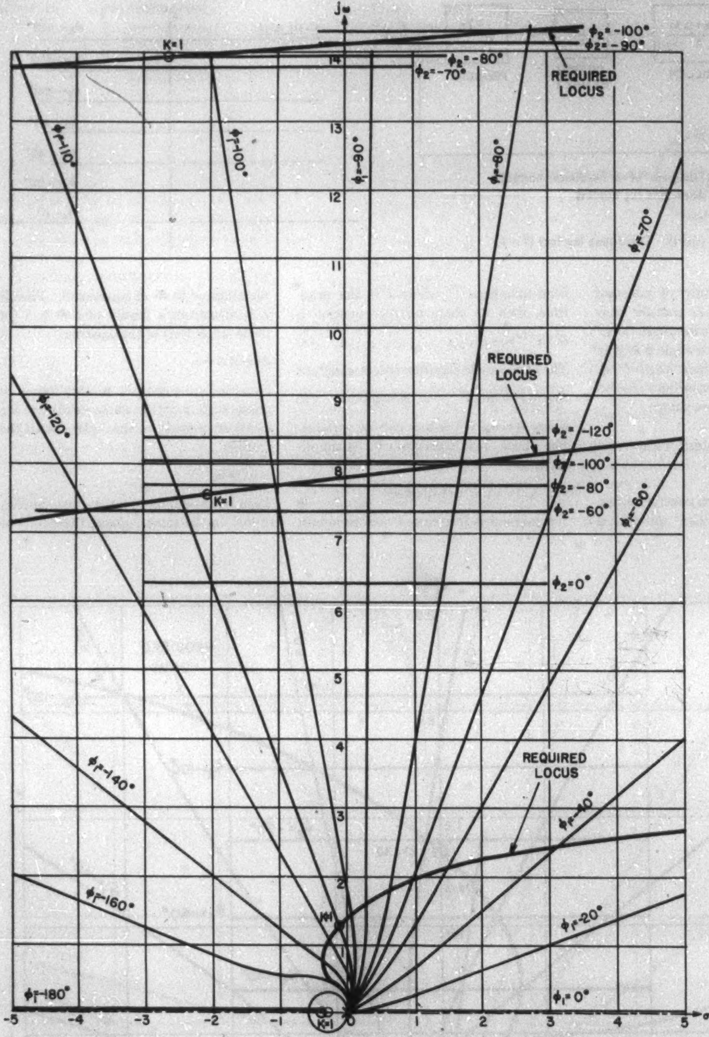


Figure 7. Three branches of the required locus with dead-time lag

intersection of the corresponding curves with the sum of phase angles ϕ_1 and ϕ_2 equal to ±180 degrees are the points of the locus required. The curve drawn through all these points of intersection is the required locus. It is the required locus because each point on it satisfies equation 6. The determination of the gain can be carried out then by the criterion used in the root-locus method. The magnitude equation is

$$K=\frac{1}{|G(\sigma+j\omega)\epsilon^{-Ts}|} \qquad (7)$$

which will give the gain K of any point on the required locus.

To illustrate and verify the method, an example is selected with the following transfer function

$$G(s)=\frac{(s+0.3)}{s^2}\,\epsilon^{-Ts}$$

For convenience, it is separated into two transfer functions

$$G_1(s)=\frac{(s+0.3)}{s^2}$$

$$G_2(s)=\epsilon^{-Ts}$$

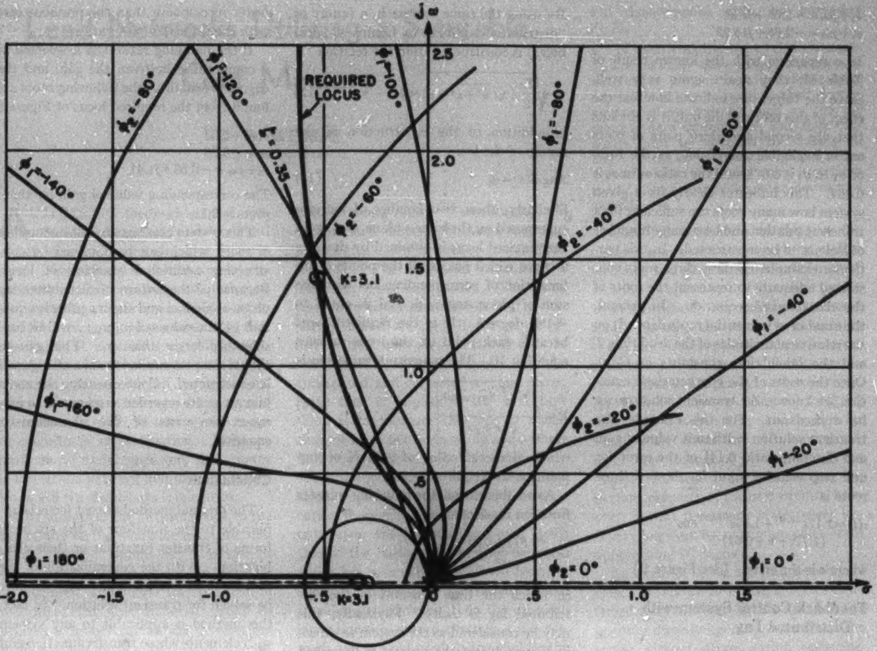


Figure 8. Construction of the required locus with distributed lag

where

$G(s)=G_1(s)G_2(s)$

and T is assumed to be unity. Physically, it may be considered as a control system with a proportional-plus-integral controller and an output member acting as an integrator and a process represented by a dead-time lag. The block diagram is shown in Figure 4. This problem was solved by Callender, Hartree, and Porter[7] in 1934. Table II is the result from that paper. It will be shown that the results of this paper agree very well with this known one.

Table II. Roots of Characteristic Equation

Order of harmonics (n+1)		
Fundamental	−0.4128	0
1	−0.1719	1.1698
2	−2.0671	7.5496
3	−2.6555	13.9280
4	−3.0216	20.2582
5	−3.2885	26.5693

By following the previous procedure, a family of root loci of the following equation is constructed

$$\text{Ang}\, G_1(s)=\phi_1$$

This family of root loci has been shown in Figure 3. Next, another family of dead-time lag loci is constructed from the relation

$$-\omega=\phi_2$$

This family of loci is shown in Figure 5. Finally, these two families of loci are superposed on the same s-plane. The required locus is the points of intersection of every two loci with the sum of ϕ_1 and ϕ_2 equal to ±180 degrees as shown in Figure 6. The points of the required locus in the upper half plane are those with a phase angle of −180 degrees and those in the lower half plane are those with a phase angle of +180 degrees. Since the locus is symmetrical with respect to the real axis, only those in the upper half plane are shown.

The required roots are the points where the value of gain K is equal to one. Three roots are found from Figure 6.

$$\sigma_0=-0.41$$

$$\sigma_1+j\omega_1=-0.17+j1.17$$

which show that they agree very well with the first three known roots of Table II. Furthermore, the value of damping ratio of the predominate pair of complex roots at unity value of gain is found to be 0.14 which is usually regarded as too oscillatory. Other values of damping ratios together with their corresponding values of gain are listed as follows

$\zeta=0.40$ at $K=0.69$
$\zeta=0.45$ at $K=0.56$
$\zeta=0.51$ at $K=0.54$

where ζ is the damping ratio. This easy way of choosing a desirable damping ratio is one of the advantages of root-locus method. The "gain limit," which is the maximum gain below which the system can have damped oscillation, can be found readily. This is the value of gain at the point of intersection of the required locus and the imaginary axis. This value is found to be 1.32.

This given transfer function has an infinite number of roots, which can be obtained by constructing additional branches of the required locus. Figure 7 shows the second and the third branches, as well as the fundamental branch. Again at the points which give the unity value of gain, we found the second and the third pairs

$\sigma_2+j\omega_2=-2.06+j7.56$
$\sigma_3+j\omega_3=-2.65+j13.92$

In comparison with the known result of Table II, they again agree very well. Since the values of σ indicate how fast the effect of this root will die out, it is obvious that the second and third pairs of roots are of negligible magnitude, as the ratio of σ_1 to σ_2 is 0.078, and the ratio σ_1 to σ_3 is 0.064. This indicates clearly for a given system how many roots are sufficient for a transient solution and how many branches of loci are to be constructed. In this particular example, the first three roots considered adequate to represent the roots of the characteristic equation. In general, the number of influential roots depends on the relative magnitudes of the dead time T and the other time constants in $G_1(s)$. Once the roots of the characteristic equation are known, the transient solution can be worked out. For this example, the transient solution (with unit value of gain and damping ratio 0.14) of the error for unit step velocity input for the first three roots is

$$e(t)=1.17\epsilon^{-0.41t}+1.35\epsilon^{-0.17t}\cos(1.17t+\pi+0.61)$$

where e is the error. (See Figure 1.)

Feedback Control System with Distributed Lag

If it is given a feedback control system with a transfer function $G(s)$ and with a distributed lag $\epsilon^{\sqrt{T_ds}}$, where T_d is the time constant, the characteristic equation is

$$G(s)\epsilon^{-\sqrt{T_ds}}=-1 \tag{8}$$

The phase-angle equation is accordingly

$$\text{Ang}\,G(s)+\text{Ang}\,\epsilon^{-\sqrt{T_ds}}=\pm 180 \text{ degrees} \tag{9}$$

The expression $\sqrt{T_ds}$ (where $s=\sigma+j\omega$) can be simplified as

$$\sqrt{T_dS}=\pm\sqrt{\frac{T_d}{2}}\left\{\sqrt{\sqrt{\sigma^2+\omega^2}+\sigma}+j\sqrt{\sqrt{\sigma^2+\omega^2}-\sigma}\right\} \tag{10}$$

In practical problems, only the positive sign has significance. By using the positive sign and substituting equation 10 into 9, the result after simplification is

$$\text{Ang}\,G(s)-\sqrt{\frac{T_d}{2}\left(\sqrt{\sigma^2+\omega^2}-\sigma\right)}=\pm 180 \text{ degrees} \tag{11}$$

By using the same approach, a family of "distributed-lag loci" (a family of parabolas) is constructed from the relation

$$-\sqrt{\frac{T_d}{2}\left(\sqrt{\sigma^2+\omega^2}-\sigma\right)}=\phi_2 \tag{12}$$

in addition to the construction of the family of the loci

$$\text{Ang}\,G(s)=\phi_1 \tag{13}$$

Similarly, these two families of loci are superposed on the same s-plane, and again the required locus is obtained by drawing a curve which passes all the points of intersection of corresponding loci with the sum of phase angles ϕ_1 and ϕ_2 equal to ±180 degrees. It is the required locus because each point on the locus satisfies equation 10. The magnitude equation is

$$K=\left|\frac{\epsilon^{\sqrt{\frac{T_d}{2}\left(\sqrt{\sigma^2+\omega^2}+\sigma\right)}}}{G(\sigma+j\omega)}\right| \tag{14}$$

which gives the value of gain K of any point on the required locus.

As an illustration, the following transfer function is selected

$$G(s)=\frac{K(s+0.3)}{s^2}\epsilon^{-\sqrt{T_ds}}$$

in which the time constant of the distributed lag is unity. Physically, this may be considered as the system as shown in Figure 4, but the process is described by a distributed lag instead of a dead-time lag. By choosing such a transfer function, the results of such a system with a dead-time lag may be compared with that with a distributed lag of the same unity time constant.

The phase-angle loci of distributed lag, which is a set of parabolas, are shown in Figure 8 (upper half plane only). The required locus, obtained by superposition of the phase-angle loci of the distributed-lag on those of Figure 3, is also shown in Figure 8. It is shown there that no damping ratio larger than 0.42 and smaller than unity can be obtained from the locus. If this damping ratio 0.42 is chosen, the value of gain K is 1.44. If the damping ratio 0.14 is chosen as the previous case of dead-time lag, the corresponding gain K is 10.6 as compared to the one of the previous case. Furthermore, the gain limit in this case is 19.6 as compared to 1.32 of the previous one. From these results as well as the shape of the locus, it may be concluded that this system is faster in response than the previous case with the same damping ratio.

If the damping ratio 0.35 is selected as a compromise between the gain and the degree of stability, the following roots are found from the required locus of Figure 8

$\sigma_1=-0.33$
$\sigma_2=-3.20$
$\sigma_3+j\omega_3=-0.53+j1.41$

The corresponding value of gain for these roots is 3.1.

This system also has an infinite number of roots, which can be obtained by constructing additional branches of locus. Because of the nature of distributed-lag phase-angle loci and the magnitude equation 14, the values of $\sigma_4, \sigma_5, \ldots$ will be considerably larger than σ_3. This can be shown if a plot similar to that of Figure 7 is constructed. Consequently, the roots just given are regarded as adequate to represent the roots of the characteristic equation.

Conclusion

The original method of root locus is applicable to the first four of the six basic forms of transfer functions or their combination. With the generalization of the root locus, all the six basic forms can now be solved for transient solution. In fact, this method is applicable to any system with elements whose transfer function and corresponding phase-angle loci are obtainable. The dead-time lag and the distributed lag are but two cases which have been used as illustrations.

References

1. PRINCIPLES OF SERVOMECHANISMS, **G. S. Brown, D. P. Campbell.** John Wiley and Sons, Inc., New York, N. Y., 1948.

2. CONTRIBUTIONS TO THE THEORY OF AUTOMATIC CONTROLLERS AND FOLLOWERS, **D. G. Prinz.** *Journal of Scientific Instruments* (London, England), April 1944.

3. AUTOMATIC CONTROL, A SURVEY OF CRITERIA FOR STABILITY, **D. G. Prinz.** *Proceedings*, Institution of Chemical Engineers (London, England) volume 27, 1945.

4. THEORETICAL FOUNDATIONS OF THE AUTOMATIC REGULATION OF TEMPERATURE, **A. Ivanoff.** *Journal*, Institute of Fuel (London, England), volume 7, February 1934.

5. THE DYNAMICS OF AUTOMATIC CONTROLS (book), **Oldenbourg, Sartorius.** Translated by Dr. H. L. Mason. *American Society of Mechanical Engineers* (New York, N. Y.), February 1948.

6. CONTROL SYSTEM SYNTHESIS BY ROOT LOCUS METHOD, **Walter R. Evans.** *AIEE Transactions*, volume 69, part I, 1950, pages 66–69.

7. TIME LAG IN CONTROL SYSTEM, **Callender, Hartree, Porter.** *Philosophical Transactions of the Royal Society of London* (London, England), volume 235 1934.

No Discussion

Lead Networks Utilizing a Saturable Core Memory

D. G. SCORGIE
ASSOCIATE MEMBER AIEE

AUTOMATIC feedback control systems have been deeply explored and widely discussed throughout the literature. This paper will discuss special circuits which are peculiarly adaptable to such systems. In particular, if the system is of the type employing a modulated carrier signal, it is possible to nullify completely the effect of a simple time lag by inserting nonlinear circuit configurations which are presented here. The circuits are uniformly effective over a wide range of magnitudes of modulation. They are particularly useful since very little power is lost in the network and low impedance loads will not disturb their operation.

The success of the circuits is due to the use of any of the available magnetic materials with "square-loop" hysteresis characteristics. R. A. Ramey in his paper on magnetic amplifiers[1] was the first to recognize that this material could be utilized to its best advantage by resetting it on one-half cycle with a small amount of power, then using it on the next half-cycle to control a large amount of power.

The new circuits given here utilize the half-cycle response concept as a memory and as a power conserving expedient, and at the same time perform the functions of a lead network filter.

Background

Among the methods of stabilizing feedback control systems, the insertion of a lead network is one of the most common. The designation lead network refers to the circuit response when a sinusoidal input of a particular frequency is applied. Throughout some important band of frequencies, the output phase of the network must lead the input. When inserted in cascade with other elements of the control loop, the lead network at least partially compensates for their lagging characteristic in that frequency range. If one is lucky, the concomitant effects upon magnitude and phase outside that frequency range will be tolerable. The term lead network is a frequency plane designation. The terms derivative circuit, rate circuit, and anticipatory circuit are used if the same effect in the time plane is described. It would be ideal if future changes could be anticipated and corrected as they occur. Since that is impossible the next best thing is to measure the rate at which changes have just been made and assume that rate will most probably continue.

For a-c systems, in which the control signal appears as the modulation upon a carrier frequency, the phase-lead action is only upon the modulation signal itself. One of the basic forms of lead network, which has been called the class-I form,[2] is shown in Figure 1. First, the modulated signal is demodulated, then the demodulated signal is passed through a d-c network and is then used to remodulate the carrier. The demodulated signal is operated on by the circuit of Figure 1 according to the following relationship

$$\frac{e_o}{e_i}=K_1\left(\frac{1+j\omega T_a}{1+j\omega T_a K_1}\right) \tag{1}$$

where

$$T_a=R_1C$$

$$K_1=\frac{R_2}{R_1+R_2}$$

For frequencies below $(K_1T_a)^{-1}$, equation 1 approximates the ideal "proportional-plus-derivative" transfer function,[3] $(1+j\omega T_a)$.

In a-c feedback control systems, class-I lead networks have the following limitations:

1. The modulator and demodulator introduce extraneous noise into the system.
2. The filter circuits of the modulator and demodulator reduce the positive phase shift produced by the lead network.
3. The maximum signal phase shift is given by

$$\sin^{-1}\left(\frac{R_1}{R_1+2R_2}\right)$$

When $R_1/R_2=10$, this maximum value is about 56 degrees. The maximum phase shift can be increased, at a sacrifice of gain and signal-to-noise ratio, by increasing R_1/R_2.

4. In servo work the desired time constants T_a are normally quite large. Nature does not readily provide us with reliable large capacitors unless we resort to their electromechanical equivalents, such as the d-c motor. As a consequence, long time constants must be obtained by making R_1 large. If the lead network is inserted before a point where the driving point impedance is low, a buffer amplifier must be inserted as well.

Work has been done to improve the characteristics of this network by adding feedback and electromechanical elements[2] in order to alleviate the first three conditions mentioned. As compared to class-I lead networks, the primary advantages of the new circuits discussed here lie in the elimination of the fourth difficulty and in their relative simplicity and reliability. As in class-I networks, the new networks must have a large voltage attenuation for steady-state conditions. A correspondingly large amount of power is lost in the class-I networks. In contrast an important property of the new circuits, however, is that the steady-state power loss is very low even for maximum phase leads approaching 90 degrees.

An example of an a-c lead network of class II is shown in Figure 2. This is referred to as the "parallel-T" network and operates directly upon the a-c modulation signal without demodulation occurring. It is difficult to design the circuit to get sufficient phase lead without attenuating the signal too greatly. In addition, the parallel-T is quite sensitive to changes in carrier frequency. This is particularly true on higher frequency systems because the absolute change of frequency, rather than the percentage change, affects the circuit operation.

Subject to the limitations previously discussed, the networks shown in Figures 1 and 2 have the transfer function given by equation 1. The simplest form of time lag element has the transfer function $K_2[1/(1+j\omega T_a)]$. When cascaded with each other, the over-all transfer function is their product $K_1K_2\ [1/(1+j\omega K_1T_a)]$. The lead network has the effect of attenuating the over-all signal by the factor K_1 and decreasing the time lag by K_1.

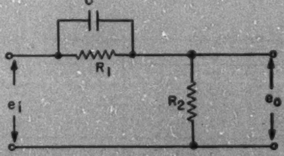


Figure 1. Class I lead network

Paper 52-272, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13-17, 1952. Manuscript submitted June 16, 1952; made available for printing August 8, 1952.

D. G. SCORGIE is with the Naval Research Laboratory, Washington, D. C.

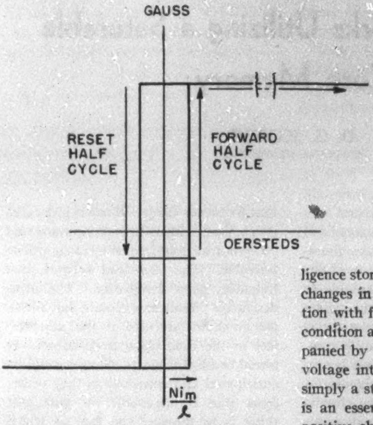


Figure 3. Hysteresis loop of grain-oriented core material

Square-Loop Toroid as a Circuit Element

Wound toroids using core materials with a square hysteresis loop have characteristics which depart so radically from those of the usual iron-cored coil that they merit very special attention. A hysteresis loop, such as that shown in Figure 3, can be obtained by properly orienting the grain structure and relieving the internal stresses in certain alloys.[4] The process results in an extremely high maximum permeability; that is, a high ratio of flux density to field intensity on the normal magnetization curve. Thinking in strictly linear terms the conclusion is reached that a coil will have a high inductance if wound on a core whose maximum permeability is high. This is not true if at the same time the retentivity of the core material is high.

An inductance which appears in linear circuit theory has certain essential characteristics which distinguish it. The pragmatist views an inductor as a phase shifting device, a high frequency attenuator, or perhaps as a time delay device. All three views are, in fact, equivalent and interdependent in linear theory. This man, testing a wound square-loop toroid, finds that when various sinusoidal voltages are applied to the winding, the current will be a square wave whose fundamental component is in phase with the applied voltage rather than lagging 90 degrees behind. The magnitude of the current will be nearly independent of the magnitude of input over wide ranges and nearly independent of frequency. If anything, the current increases with increasing frequency due to eddy current effects in the core. The time-delay aspect is present, although in a completely different form. When a step function voltage is applied, a step of output current occurs an interval of time later which depends upon the magnitude of the input and the previous history of the core. In any event, the pragmatist will conclude from his tests that his wound toroid does not act like a law-abiding inductor.

The theorist might think of an inductor as an energy storage device. During the time current through the coil is increasing, energy given by $\int EIdt$ is being drawn from the source, part of which is being stored in the magnetic field. He can measure the power, if any, drawn from the source, while the current is being reduced to zero, make allowance for heat dissipated in the coil resistance, and then easily calculate how much energy was stored in the inductance and returned to the circuit. At the point of maximum current, the energy ultimately returned to the circuit is stored in the form of internal stresses within the core material. In the grain-oriented core materials these internal stresses are absent. If high currents flow through windings on these cores and the circuit is opened by a switch, the arc will be small and due only to the small energy stored in the equivalent air-core inductor. The theorist also will conclude that the coil wound on a grain-oriented core has negligible inductance.

How then must we view the grain-oriented core? Its primary function is as an information storage device or memory. Additions or subtractions from the intelligence stored in the core take the form of changes in the degree of magnetic induction with full saturation as the reference condition and must necessarily be accompanied by the presence of a proportional voltage integral across the coil. This is simply a statement of Faraday's law. It is an essential requirement, before any positive change of induction ΔB can occur, that a positive magnetizing current i_m flow through the winding. Correspondingly, $-i_m$ must flow before any negative change in induction can take place. The latter statements are simply a verbal picture of what is meant by "square-loop" hysteresis. Viewed from the external circuit, the following steps describe the method of utilization of the square-loop core:

1. All information can be cleared from the core by applying for a sufficient length of time a voltage greater than that necessary to force i_m through the winding and series resistance. The total resistance in series with the coil and the voltage source must be low enough to permit the current to reach the value i_m, usually at a negligibly small voltage.
2. When information is to be stored, the polarity of the voltage source becomes negative. The term reset is applied to the process of storing information, starting from a fully saturated core condition. Since the reset voltage $-E_r$, across the coil is proportional to $(dB)/(dt)$, information is presented to the coil as a controlled quantity $\int E_r dt$ which uniquely determines a change of induction $-\Delta B$. Thus a level of induction is established which will remain stored until needed.

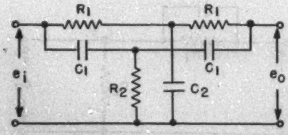


Figure 2. Class II lead network

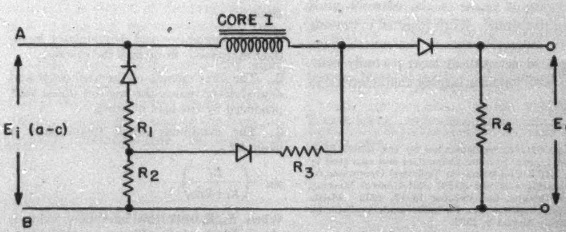


Figure 4. Basic lead network, half-wave d-c output

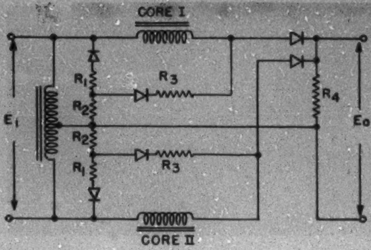


Figure 5 (left). Lead network, full-wave d-c output

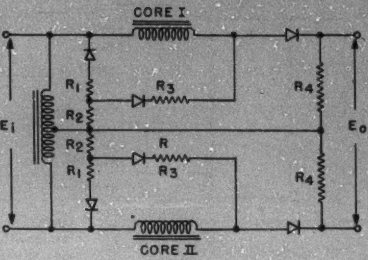


Figure 6 (right). Lead network, full-wave a-c output

3. The information is utilized by applying a known integral of forward voltage $\int +E_f dt$ in general larger than the preceding reset voltage integral. At the outset, the core winding appears as a large resistor, allowing only $+i_m$ to pass, no matter what the instantaneous value of the forward voltage. As soon as the core saturates, the apparent resistance of its coil disappears and a large current can flow to the external circuit. This current surge is a function of the stored information.

Operation of the New Lead Networks

Differentiation can be done by a cascade of magnetic amplifiers as has been shown in a Naval Research Laboratory patent application by R. A. Ramey. If proportional-plus-derivative action from a single core is to be obtained, the circuit of Figure 4 will perform that function as explained in the following. Core I is a wound toroid of square-loop material. An alternating voltage source is applied at terminals A-B and the output is a half-wave pulsating direct current which appears across resistor R_4. The input can have any wave shape whatsoever. One can infer from the preceding section that only the half-cycle voltage integral is important.

Each half-cycle of input where terminal A is positive relative to terminal B is called a forward half-cycle. The effective circuit during a forward half-cycle is core I and the resistor R_4 in series. No current flows through R_1, R_2, and R_3. The voltage E_{ac} will appear either all across core I or all across R_4, depending upon whether or not the core remains unsaturated.

Half-cycles of input where B is positive relative to A are reset half-cycles, during which R_4 is inactive. The voltage applied to core I, during reset, is determined by R_1, R_2, and R_3, the latter being primarily to compensate for differences in forward resistance of the various rectifiers in the circuit at low voltages.

The circuit of Figure 4 is nonlinear in the sense that it contains a saturable core device and that it introduces noise which has frequency components higher than the carrier frequency. However, it is linear in the more important sense that the average value of the output and modulation of the input are related by a unique transfer function $1+Ts$ where s is the Laplace operator, and T is an adjustable constant. The term Ts arises where there is a component in the output proportional to the derivative of the input. The transfer function is valid for all frequencies up to the carrier frequency independent of the value of T.

The definition of the derivative of the function $f(t)$ with respect to time involves a knowledge of the function at pairs of distinct times. Specifically, the derivative of $f(t)$ is given by

$$\frac{df}{dt} = \lim_{(t_2-t_1)\to 0}\left(\frac{f(t_2)-f(t_1)}{t_2-t_1}\right)$$

If the value of $f(t)$ only at time t_2 is known, there is no clue as to what corrections must be made in the immediate future. If, however, we can remember what the value was a moment earlier, at time t_1, (that is, if the derivative is known) we can usually anticipate fairly well what is coming. The philosophy involved in the circuit conforms with this definition. One can remember or store information concerning the magnitude of half-cycle of the carrier frequency, then can compare it to the magnitude of the next half-cycle. If there is a difference in the voltage integrals of these two half-cycles, the circuit will sense the fact and apply an extra large or small pulse to the output terminals.

We will assume that at the beginning of the reset half-cycle, the core is saturated (at the upper end of Figure 3). The integral of the input voltage during the reset half-cycle is $\int_r E_{ac}(t)dt$ and the voltage integral impressed upon the core is $\gamma \int_r E_{ac}(t)dt$, where $\gamma = R_1/(R_1+R_2)$. The flux level therefore will change $N\gamma \times 10^8 \int_r E_{ac}(t)dt$ where N is the number of turns upon the core. The input might be some function such as $E_{ac}(t) = \mathbf{E}^*_{ac}(t) \sin \omega t$, where the coefficient on the right can be any function of time. On the forward half-cycle the magnitude will, in general, be different from its value on the previous reset half-cycle, and the area of the forward half-cycle will be designated $\int_f E_{ac}\, dt$.

As long as the core does not saturate on the forward half-cycle it absorbs the full input voltage. After saturation, the full input voltage is impressed across R_4. Therefore, the integral of voltage across R_4

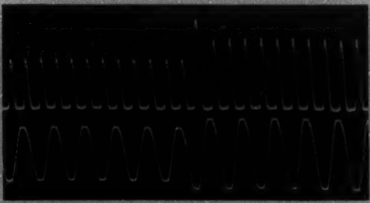


Figure 7 (left). Response of lead network to positive step input

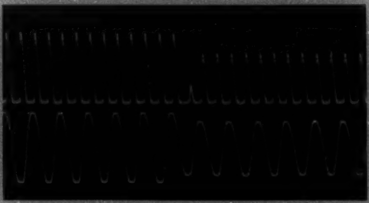


Figure 8 (right). Response of lead network to negative step input

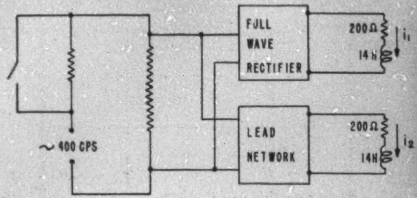


Figure 9 (above) Test circuit for evaluating effect of lead network

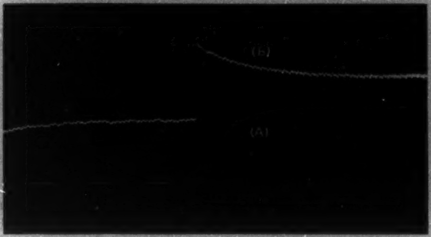


Figure 11. Build-up of current through inductance-reactance circuit

(A) i_1 of Figure 9

(B) i_2 of Figure 9, time constant too large

during the forward half-cycle is

$$\int_f E_0 dt=\int_f E_{ac}(t)dt-\gamma\int_r E_{ac}(t)dt \qquad (2)$$

Adding and subtracting $\gamma\int_f E_{ac}dt$ to the right half of equation 2 we get

$$\int_f E_0 dt=(1-\gamma)\left\{\int_f E_{ac}dt+\left(\frac{\gamma}{1-\gamma}\right)\left[\int_f E_{ac}dt-\int_r E_{ac}dt\right]\right\} \qquad (3)$$

If the input magnitude is not varying, the term in the square brackets on the right will be zero and the output simply will be proportional to the input. Equation 3 can be written in more meaningful symbols after first dividing through by the half period time Δt. The average output on the forward half-cycle is defined by $G_0\equiv(\int_f E_0 dt)/\Delta t$ and the average input during the same interval by $G_a(\equiv\int_f E_{ac}dt)/\Delta t$. The quantity in the square brackets of equation 3, when divided by Δt is the change in two successive half period averages of input and is designated ΔG_a. We shall multiply and divide ΔG_a by Δt and get $(\pi/\omega)(\Delta G_a/\Delta t)$. By definition $T=\gamma/(1-\gamma)\ \pi/\omega$ and the second term on the right of equation 3 becomes $T(\Delta G_a/\Delta t)$. If the modulation frequency is low, relative to that of the carrier, the ratio of increments becomes very nearly a derivative and we shall assume that $(\Delta G_a)/\Delta t=(dG_a)/dt$. Equation 3 then becomes

$$G_0=(1-\gamma)\left(G_a+T\frac{dG_a}{dt}\right) \qquad (4)$$

The constant γ can be varied from zero to unity; correspondingly, the time constant T varies from zero to infinity. It should be noted, however, that T is inversely proportional to the frequency of the carrier.

Equation 4 shows the proportional-plus-derivative relationship between the input voltage and the average output. The only assumption necessary was that the frequency of modulation of the input be lower than the carrier frequency.

If we wish to raise the frequency of the lowest noise component in the output, a full-wave circuit can be used, see Figure 5. If, instead, an a-c output is desired, one circuit configuration is that of Figure 6.

Figures 7 and 8 show responses of the full-wave d-c circuit, see Figure 5, to step function changes of a-c input. The extra large and extra small pulses reflect, respectively the positive and negative derivatives in the input steps.

The full-wave circuit has the added advantage that it eliminates ambiguity and some of the "dead-time" in the operation, when sudden changes of input occur. The half-wave circuit will deliver no derivative term whatsoever if the input change occurs as a step function at the beginning of a reset half-cycle. The full-wave circuit will always deliver the full voltage integral corresponding to a step change, although the derivative portion may be shared between the next two pulses depending upon where in the half-cycle the step occurred.

Practical Considerations

To illustrate the effectiveness of the new circuit in eliminating a time lag, the build-up of current in two identical inductance-reactance circuits was studied. Each circuit consisted of a 14-henry choke and a 200-ohm resistor in series. In each case a step function change of alternating current was applied by means of the test circuit shown in Figure 9. The first circuit was driven through a full-wave rectifier and the second driven through the lead network, Figure 5. Curve A, in Figure 10, shows the exponential current rise through the first inductance as contrasted with the very rapid rise, curve B, through the second inductance. The presence of greater ripple in the latter case is due to the relatively great higher harmonic content of the pulsed wave. By simply adjusting the potentiometers R_1+

Figure 10 (below). Build-up of current through inductance-resistance circuit

(A) i_1 of Figure 9

(B) i_2 of Figure 9, perfect adjustment of lead network

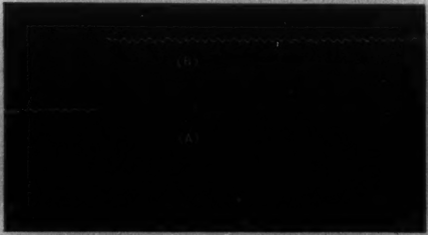
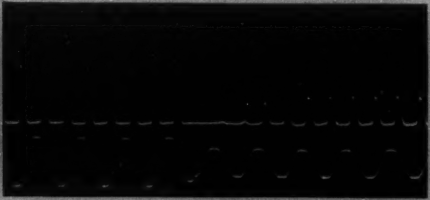


Figure 12 (below). Response of lead network to large negative step input



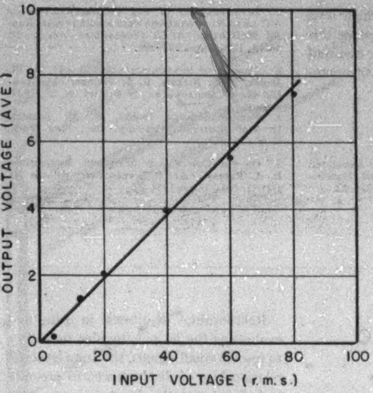


Figure 13. Steady-state response characteristic of lead network

R_2 the current through the inductance could be made to overshoot, see Figure 11, undershoot, or as shown in curve *B*, Figure 10, could be made to follow the step function input practically instantaneously. The potentiometers adjust the size of the initial pulse due to the derivative effect. If the pulse is too large, the overshoot in the build-up is virtually instantaneous and the current then decays exponentially to its steady-state value with the time constant L/R.

The lead networks are sensitive to the frequency of the carrier as has been shown. This is a reflection of the fact that for a given load resistor and a given rectified average input, more energy is available to supply to the load in a single half-cycle of the lower frequency carrier. The sensitivity to shift of carrier frequency is not so great as that of the parallel-T lead network. Furthermore, no steady-state output drift will result from a carrier frequency shift. Only the time constant is affected. The change of time constant is inversely proportional to frequency and hence is no more sensitive at high frequencies than at low.

Normally the circuits discussed here are intended to be used with modulation frequencies far less than the carrier frequency. If this mode of operation is not used; that is, if a step input is applied, the analysis passing from equation 3 to equation 4 is not valid. If, furthermore, a very large step input occurs, equation 3 is not necessarily valid because a type of overloading effect, peculiar to very large negative steps, destroys its validity. Figure 8 illustrates that the output pulse immediately following a negative step of input is very nearly zero. If the step had been any larger, the pulse would have disappeared altogether as it can never itself be negative. Instead, the amount by which it failed to be negative is subtracted from the immediately succeeding pulse or pulses as shown in Figure 12. This behavior is explained as follows. Normally, core I saturates near the end of every forward half-cycle. However, if the input magnitude is reduced by a step function at the beginning of a forward half-cycle, the full forward voltage integral may not be enough to saturate the core. The remaining difference in flux represents the missing negative pulse. Each subsequent cycle will move the core closer to saturation until finally an output pulse is restored.

Linearity of the proportional part of the output is important in some applications. Here the various resistances, particularly the forward resistances of the rectifiers, become important. The resistor R_3 is in series with the crystal rectifier in the various reset circuits to equalize the rectifier resistances at low applied voltages. The other rectifiers are selenium and have very high forward resistances at low currents. With R_3 present, a typical steady-state response characteristic of the half-wave circuit, Figure 4, is shown in Figure 13. Again it is emphasized that the steady-state curve shown here will not be influenced by frequency changes of the carrier. Where linearity of the steady-state response is extremely important, a selenium rectifier can be inserted in place of R_3 and the crystal rectifier. Also, a rectifier can be inserted in series with R_2. These expedients are designed to equalize forward rectifier voltages at low input voltages but should be used only as a last resort, for unless fairly large input voltages are used, the temperature effects in the various rectifiers cause drifting of the steady-state output. An alternative approach, which is often better, is to lower the forward resistance of the various selenium rectifiers at low applied voltage by inserting in parallel with each a series combination of crystal rectifier and resistor.

It has been stated earlier that a primary function of the new circuit was to replace class I and class II lead networks when the following stage must be a low impedance device. For large time constants the class I and II networks must have either a large resistor or capacitor. Nature does not readily furnish us with reliable large capacitors so that the high value of resistor must be used. The nonlinear lead networks on the other hand, perform the function of the very large capacitors, while allowing the series resistance to be very small.

The time constant, for a given carrier frequency, is determined by the resistors R_1 and R_2, Figure 4, through which the load current never flows. The test shown in Figure 9 illustrated a result with a random set of parameters; and the best job that could be done with a class I lead network, would have required at least a 200-microfarad capacitor.

The power conserving aspect of the new circuits is often of prime importance. In those instances it is well to consider that resistor R_4, Figure 4, need not be linear. Its only function is to provide a path for the magnetizing current while core I is unsaturated. After saturation R_4 could just as well be nonlinear, such as a blocking rectifier. To let magnetizing current i_m flow prior to saturation a current source would have to provide forward current through this rectifier.[5] If the load is an inductance in parallel with the blocking rectifier R_4 the combination furnishes the low impedance path for i_m.

If the magnetic materials being used have reasonably high retentivity but not sufficiently high to give the square-loop appearance of Figure 3, the ideas outlined in this paper can still be utilized, but not quite so simply. Additional elements must be inserted to nullify the effect of the energy stored in such a core at peak flux density.

Summary

A new philosophy of obtaining phase lead for a modulated carrier signal has been presented with several examples of the implementing circuitry. It has been demonstrated that the new circuits with very small power loss are capable of nullifying the effect of a cascaded simple time-lag element completely, provided that the time lag is of such a type as to be primarily sensitive to the rectified average of the applied voltage. The heart of the device is a wound toroid of grain-oriented iron with the square-loop hysteresis characteristic. Material of this type is capable of retaining stored information, which can be compared at a later time with other information with which it is presented. It is seen that in a sense this material can reliably provide the essential performance of a capacitor of 1,000 microfarads or more.

References

1. On the Mechanics of Magnetic Amplifier Operation, **Robert A. Ramey.** *AIEE Transactions*, volume 70, part II, 1951, pages 1214–23.

2. Improvements in the Characteristics of A-C Lead Networks for Servomechanisms, **Donald McDonald.** *AIEE Transactions*, volume 69, part I, 1950, pages 293–300.

3. Theory of Servomechanisms (book), **H. M. James, N. B. Nichols, R. S. Phillips,** McGraw-Hill Book Company, Inc., New York, N. Y., 1947.

4. Ferromagnetism (book), **R. M. Bozarth,** D. Van Nostrand Company, Inc., New York, N. Y., 1951.

5. On the Control of Magnetic Amplifiers. **R. A. Ramey.** *AIEE Transactions*, volume 70, part II, 1951, pages 1214–23.

No Discussion

Vacuum-Tube Techniques Applied to a Hydraulic Amplifier

J. A. BARING
ASSOCIATE MEMBER AIEE

SCIENCE endeavors to describe large numbers of situations observed in nature by the simplest and most general theory. The more generally applicable a theory, the more useful and highly regarded it becomes.

The analogies which can be drawn between the behavior of seemingly unrelated devices have great appeal since a general approach is thus capable of describing the basic properties of two or more systems. The analogy frequently drawn between the dynamics of mechanical, electric, and acoustical systems is an example.

This paper deals with the striking similarity that can be observed between the behavior of a hydraulic amplifying device known as the jet pipe and the behavior of the triode vacuum tube. When thought of as a "hydraulic triode," the jet pipe and its associated hydraulic circuit can be treated in a manner well known to electronic circuit designers. Once the hydraulic characteristic curves are established, such concepts as the quiescent operating point, the internal impedance, the equivalent circuit, and the efficiency and power output can be derived.

Description of the Jet Pipe

The jet pipe is a mechanical hydraulic amplifier. It is functionally similar to the conventional 4-way pilot valve in that fluid is directed to one of two output ports in response to a mechanical displacement. A sketch of the unit is shown in Figure 1.

The long tube or pipe is pivoted about an axis as shown by an amount θ. The oil flow Q emerges from a nozzle-like tip at a velocity of roughly 1,000 inches per second and impinges on two adjacent pickup holes at 1 and 2. A pressure differential ΔP builds up across the holes, dependent on θ and the load flow S. θ is the angular rotation of the jet pipe measured from the center position. Action is symmetrical about the $\theta=0$ position but opposite in sense. The torque required to displace the pipe can be made extremely low. No turning reaction is transmitted to the pipe from the load; however, it is subjected to low-level random noise affects of splashing oil at the tip.

Any device providing a displacement can be used as an input source.

It is usually convenient to apply force inputs which are converted to proportional displacements by the use of centering springs. For example, a diaphragm may be connected to the jet pipe by means of a push pin. When relatively light centering springs are used, output changes for as little as 10^{-4} pound-per-square-inch change in diaphragm pressure can be realized.

Relationships of interest in using and evaluating the jet pipe involve its ability to resolve small signals, the noise level, its relationships of displacement to pressure and flow, and its dynamic properties. Some of these have been dealt with by Dushkes and Cahn.[1] The principal discussion here will concern the pressure-flow characteristics and their analogy to those of the triode vacuum tube.

Hydraulic Circuit

A physical schematic of the jet pipe, pump, and load is shown in Figure 2(A). With a steady-state displacement of the jet pipe $+\theta$, the total jet pipe and pump flow Q will divide at the distributor block into a load flow S and a rejected flow R. These two will be combined at the sump and recirculated. To consider the system as a simple series circuit, the flow R is ignored and only the useful load flow S is considered. By also ignoring the pressure drops in the connecting tubing and in the pipe stem itself, it is possible to draw the series circuit of 2(B) to describe the system. The three circuit elements are the pump, producing a pressure rise independent of flow, the jet pipe, and the load.

The box representing the jet pipe produces a pressure drop in the direction of the flow S which would be realized physically, Figure 2(A), as the sum of the drops across each pick-up hole. For the sake of simplicity these two drops are lumped into one box which represents the over-all pressure-flow characteristic of this element.

In the circuit of Figure 2(B), the flow must be continuous and the pressure drops around a closed loop must add to zero.

Analysis

Some theoretical justification for the jet-pipe characteristics can be established

---

Paper 52-298, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted July 25, 1952; made available for printing September 3, 1952.

J. A. Baring is with the Askania Regulator Company, Chicago, Ill.

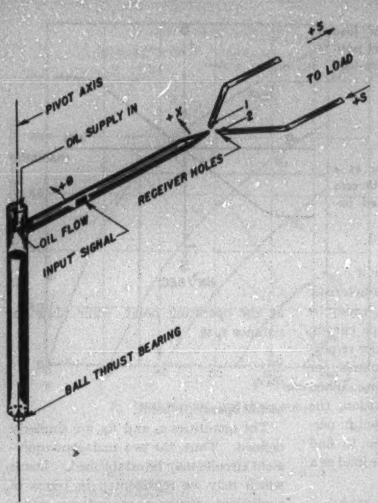


Figure 1 (left). The jet-pipe amplifier

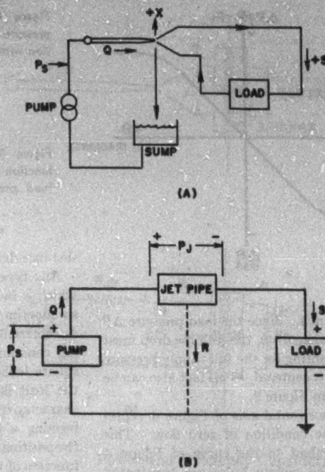


Figure 2 (right). (A) Jet-pipe amplifier as part of a hydraulic series circuit. (B) Simplified series circuit used for analysis

with the help of certain simplifying assumptions. The system can be represented as in Figure 3.

The stream and pickup holes are considered to be equal squares with sides a, for ease of manipulation. This is a fair approximation for small deflections of the jet pipe from center.

The flow from the stream divides in proportion to the deflection of the pipe from the center line between the holes. The lateral flow from the left hole is equal to the proportion of Q received minus the flow absorbed by the load. For convenience in this analysis let X, the tip motion, equal zero when the stream strikes hole 2 directly. Then $X = a/2$ when the stream is centered, that is, when $\theta = 0$.

Thus

$$Y = \frac{X}{a} Q - S \qquad (1)$$

Likewise for the right hole

$$Z = \left(\frac{a-X}{a}\right) Q - S \qquad (2)$$

The force imposed on the fluid column 1 in Figure 3 is due to the flow Y entering with velocity v_{y1} and leaving with velocity v_{y2}, which has a zero vertical component. Likewise the force on column 2 in Figure 3 is imposed by the flow Z. The force resulting from the rate of change of fluid momentum is given by $F =$ (flow) (mass density) (net velocity in the direction of force).

Thus

$$P_1 = \frac{F_1}{a^2} = \frac{Y\rho}{a^2}\left(\frac{Q}{a^2}\right) \qquad (3)$$

and

$$P_2 = \frac{F_2}{a^2} = \frac{Z\rho}{a^2}\left(\frac{Q}{a^2}\right) \qquad (4)$$

where

$P_1 =$ the pressure at 1 in figure 3.
$\rho =$ fluid mass density
$Q =$ jet flow, a constant.

Then the differential pressure across holes 1 and 2, Figure 3

$$\Delta P = P_1 - P_2 = \frac{Q\rho}{a^4}(Y - Z) \qquad (5)$$

Substituting for Y and Z from equations 1 and 2, we have

$$\Delta P = \frac{\rho Q}{a^4}\left(\frac{2Q}{a} X - Q - 2S\right) \qquad (6)$$

Equation 6 describes jet-pipe operation.

It can be seen that if the load flow S is zero, then the differential pressure ΔP increases linearly with X from zero, when $X = a/2$, $(\theta = 0)$, to the maximum of $Q^2\rho/a^4$ when $X = a$.

The experimental curve, Figure 4, has verified this linear relationship. It was obtained by dead-ending the receiver holes and measuring the differential pressure. The linearity holds in the center region.

It is further possible to measure the relationship between flow S and the displacement X for $\Delta P = 0$. Equation 6 again indicates that S is a linear function of X. This is borne out by the experimental curve, Figure 5.

X may be replaced by θ in these considerations of linearity.

Characteristic Curves

The jet pipe can be considered as a circuit element. Of primary importance is the establishment of its pressure-flow characteristics so that it may be treated analytically in hydraulic circuits. Equation 6 predicts a linear relationship between load pressure ΔP and load flow S, and for both ΔP and S as functions of X. Experimentally these curves are not quite straight lines. A typical family is shown in Figure 6. On the horizontal axis is plotted P_j, the pressure drop across the jet pipe as the series circuit element in Figure 2. On the vertical axis is the useful series flow S.

The curves represent values of constant position with a passive-type load. The supply pressure shown is 100 pounds per

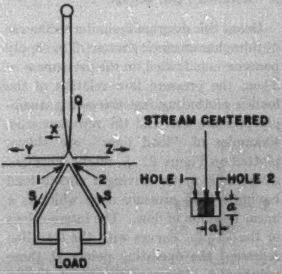


Figure 3. Jet-pipe amplifier flow diagram. Top view of stream and holes (assumed square)

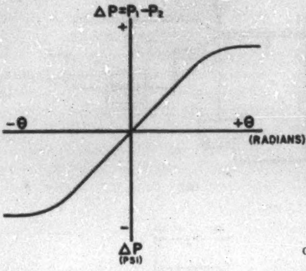


Figure 4 (left). Differential load pressure as a function of input position with zero load flow (infinite load impedance)

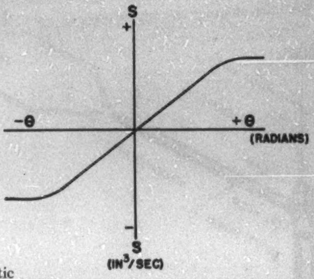


Figure 5 (right). Load flow as a function of input position with zero load pressure drop (zero load impedance)

square inch. Since the load pressure ΔP is zero when $\theta=0$, the jet-pipe drop must be accounting for the full supply pressure when it is centered. This fact also can be seen from Figure 2.

The horizontal axis of Figure 6, represents the condition of zero flow. This is equivalent to the curve of Figure 4 where the pickup holes are terminated in gauges. When $\theta=0$ the jet-pipe drop is equal to the supply pressure and no drop appears across the load. As θ is increased, more and more pressure appears across the load. Eventually, the jet-pipe drop is zero (ignoring losses) when the maximum value of θ has been reached.

In like manner, Figure 5 is repeated along the 100-pound-per-square-inch line. Since the load absorbs no pressure drop, the jet pipe must assume the full 100 pounds per square inch, regardless of θ and the flow increases with θ.

Vacuum-Tube Analogy

A strong resemblance exists between the jet-pipe curves on Figure 6 and those of the typical triode vacuum tube. Analogous quantities are:

1. Flow—plate current
2. Pressure drop—plate voltage
3. Rotation—grid voltage

Use of this diagram is similar to the use of tube characteristic curves. The supply pressure is indicated on the pressure axis. Then, the pressure flow relation of the load is plotted against the constant-supply pressure axis and the zero-flow axis. Examples of "load lines" are shown plotted on Figure 6. L_1 might represent a hydraulic motor driving a viscous load having a load pressure drop which is a linear function of flow. The intersections of the jet-pipe curves with the load line represent the operating points at those values of θ. The operating point divides the horizontal distance into the jet-pipe drop and the load drop. The vertical distance determines the series flow S.

Any type of passive load characteristic which is independent of position may be so superimposed on the jet-pipe curves. L_2 represents a hydraulic cylinder requiring constant pressure of 5 pounds per square inch to maintain motion. When the load is a function of position, the characteristic curves may be used in performing a graphical integration to find the position-time behavior of the load as a function of the jet-pipe input.

Internal Resistance

From the characteristic curve, we may define the jet-pipe internal resistance Rj_p. It is equal to the reciprocal of the slope of the characteristic curve at any operating point. On Figure 6 is shown a tangent to the θ maximum curve representing $1/Rj_p$ at that point. The dimension of Rj_p is (pressure)/flow. Rj_p is not constant over the plane, being generally smaller for lower values of flow.

The internal resistance has the character of damping. When considering the dynamics of systems including the jet pipe, the internal resistance Rj_p becomes of considerable interest. The jet pipe may be used as a pressure generator, for example. The internal resistance Rj_p may constitute the principal damping force within the circuit. Changes in Rj_p must therefore be investigated to determine their effect on the over-all dynamics of the system.

Equivalent Circuit

The tube-jet pipe analogy may be carried further by establishing similar equivalent circuits. Figure 7 shows the comparison between the basic electric and hydraulic circuit arrangements. The analogy is evident from the figure. For incremental changes about a given quiescent operating point, the electric equivalent circuit 7C is valid. μ_e, the amplification factor, is defined as

$$-\frac{\partial e_b}{\partial e_g}\bigg|_{i_b}$$

at the operating point, while plate resistance r_p is

$$\frac{\partial e_b}{\partial i_b}\bigg|_{e_g}$$

at the operating point.

The quantities μ_h and Rj_p are similarly defined. Thus, the two analogous equivalent circuits may be established. Loads which may be represented in terms of equivalent resistance, capacitance, or inertia may be connected to these equivalent circuits for analysis.

The jet-pipe circuit might prove useful in the following way. Suppose that it were of interest to determine the transfer function

$$\frac{X}{\theta}(s)$$

of the system shown in Figure 8(A).

This cylinder-spring-restriction unit may be used as a hydraulic computing device in automatic control applications. Its function is to provide a retractile feedback around an integrator to achieve proportional plus reset control.

One method of attack is to draw the equivalent jet-pipe circuit as in Figure 8(B).

Let

M = the oil and cylinder mass
B = cylinder-piston damping
K = spring constant
R = a linear hydraulic resistance representing the restriction (pressure)/flow

Assuming no compressibility and no leakage, the basic equations may be written as

$$\Delta P=rR$$

$$q=r+A\frac{dx}{dt}$$

$$A\,\Delta P=M\frac{d^2x}{dt^2}+B\frac{dx}{dt}+KX$$

$$\mu_h\theta-qRj_p=\Delta P$$

These may be combined to express the output X in terms of the input θ, as

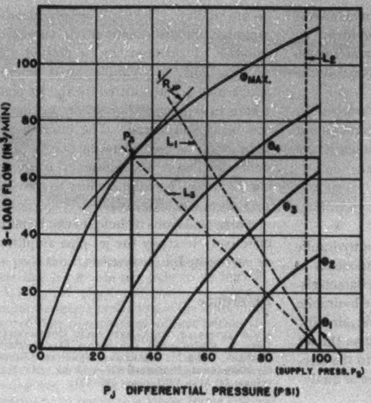


Figure 6 (left). Jet-pipe characteristic curves with sample load lines superimposed. $\theta_{\max} > \theta_1$

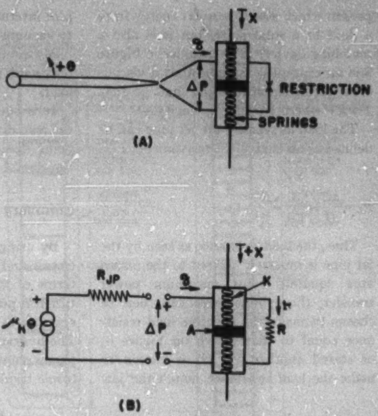


Figure 8 (right). (A) Jet-pipe amplifier as part of a hydraulic computing system. (B) Equivalent circuit of the hydraulic computing system

$$\frac{X}{\theta}(s)=\frac{\mu_h c_1}{M_s{}^2+(B+c_1ARj_p)s+K}$$

where

$$s=\frac{d}{dt} \text{ and } c_1=\frac{A}{\left(1+\frac{Rj_p}{R}\right)}$$

If M is considered to be small, then

$$\frac{X}{\theta}(s)=\frac{K_1}{\tau_1 s+1}$$

where

$$K_1=\frac{\mu_h c_1}{K}$$

and

$$\tau_1=\frac{B+c_1ARj_p}{K}$$

Thus, the transfer function has been evolved using the concept of the equivalent circuit. This device behaves, to a first approximation, like a single order lag. Its computing properties can be evaluated in the control loop on this basis. In this simple case the analytical solution might have been written directly, however, where several loops are involved the equivalent-circuit techniques can be used to advantage.

Power and Efficiency Consideration

The useful load power is equal to the instantaneous product of load pressure and flow. A graphical representation of load power for a particular operating point P_3 is shown on Figure 6. The area outlined is bounded by the load flow and the load pressure drop.

The operating point P_3 represents the point of maximum load power. As might be expected, this occurs for θ maximum; the maximum jet-pipe angular deflection. If we assume that the load is resistive in nature, the load can be plotted as shown. At P_3, the jet-pipe internal resistance Rj_p is equal to the load resistance. This can be seen in that the two lines are perpendicular at the point of intersection. It is consistent with the electrical concept of matching the load impedance to that of the output stage in order to achieve maximum power transfer.

Impedance Matching

Impedance matching may be accomplished in hydraulic output stages by the correct choice of cylinder or motor size.

Thus, the cylinder is not only a com-

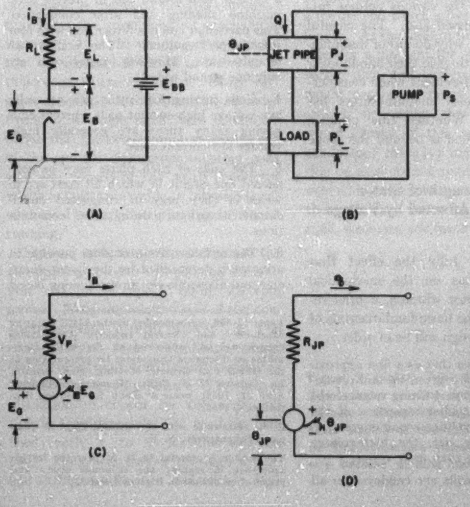


Figure 7 (left). Equivalent circuits

(A) Vacuum-tube circuit
(B) Comparable hydraulic circuit
(C) Vacuum-tube equivalent circuit
(D) Jet-pipe equivalent circuit

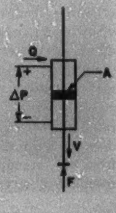


Figure 9. The hydraulic output cylinder as an impedance matching device

ponent which acts to convert energy from a fluid to a solid state, but it is also a matching device. The cylinder in Figure 9 is an example. Assume that a viscous type force is applied to the shaft so that $F=RV$ where $V=$ shaft velocity.

The load as seen by the jet pipe can be defined as an equivalent resistance

$$R_L=\frac{\Delta P}{Q}=\frac{\dfrac{RV}{A}}{VA}=\frac{R}{A^2}$$

Thus, the load resistance as seen by the jet pipe is inversely related to the piston area squared. For maximum power transfer, the cylinder area can then be chosen to make the effective load resistance equal to that shown on Figure 6; or stated another way, it is chosen to make the load resistance match the jet-pipe internal resistance. Again an analog to vacuum-tube techniques is evident.

Maximum efficiency is obtained when maximum power is being delivered to the load.

It is equal to the power absorbed by the load divided by the total pump power being dissipated. Again a similarity to the class *A* amplifier is notable.

Summary

By using the concept of the hydraulic circuit and by describing components in terms of their pressure-flow characteristics, it is possible to manipulate hydraulic systems as is done in electric circuits. The hydraulic amplifier known as the jet pipe is shown to be analogous to the electronic triode tube. Techniques of analysis are similar. The use of graphical construction on the characteristic curves is familiar. Load lines can be plotted. A jet-pipe equivalent circuit can be drawn which may be substituted for the jet pipe where incremental changes of the input are being considered. The internal resistance for a given operating point can be defined. Power and efficiency concepts can be used for load matching to provide maximum power transfer.

These analogous techniques can be used effectively to study the jet-pipe amplifier as part of hydraulic systems.

Reference

1. Analysis of Some Hydraulic Components Used in Regulators and Servo Mechanisms, **Dushkes, Cahn.** *Transactions*, American Society of Mechanical Engineers (New York, N. Y.), volume 74, May 1952, pages 595–601.

No Discussion

Voltage Rating Versus Horsepower of Synchronous and Induction Motors

C. E. MILLER
MEMBER AIEE

EVERY industrial plant of any size today has two or three utilization voltages, any one of which is available for supplying motors. The maximum primary distribution voltage within the plant buildings may be any voltage up to 15 kv, the maximum standard being 13.8 kv. Thus a choice must be made as to the most desirable voltage rating for the various motor ratings. Many factors must be considered if an intelligent selection is to be made. Each of these factors will be discussed to the end that the recommended minimum horsepower for each voltage level may be determined. This paper is concerned with the recommended minimum and not with the minimum horsepower at the various voltage ratings that it is possible to manufacture. Stated otherwise, it is not what can be built but what should be built from the standpoint of the best over-all installation for the purchaser. Likewise, just because a motor is listed in the motor manufacturers price handbook, for example a motor rated 50 horsepower (hp), 2,300 volts, does not mean that it is as good a motor from the standpoint of long life and trouble-free service as a similar motor rated 440 volts. It does indicate, however, that such a motor can be built if absolutely necessary. Maximum horsepower ratings at the various voltage ratings will be discussed also. The natural question then is: why do motor manufacturers list such ratings in their handbooks? The answer is self-evident when considering that all motor manufacturers list these ratings. Whether this practice should be continued is beyond the scope of this paper.

Mechanical Strength of Stator Windings as Affected by Voltage Rating

To appreciate fully the effect that voltage rating has on the mechanical strength of a motor winding, a brief review of some of the basic fundamentals of motor winding design will be in order.

1. It is well known that as a first approximation, a motor of a given hp and revolutions per minute (rpm) rating requires the same amount of copper regardless of the voltage. If high voltage requires going to an oversize frame size, the high-voltage motor may require even more copper.

2. Form wound coils are employed in all high-voltage (above 600 volts) motors regardless of frame size. Random windings are used in low-voltage motors; in frame 505 (125 hp, 1,800 rpm) and smaller for induction motors; and frame 945 (100 hp, 1,800 rpm) and smaller for synchronous motors.

3. Form-wound coils may cover the range from a single turn coil to approximately 20 turns. However, three turns is the usual practical minimum. In coils having few turns, each turn is made up of several strands in parallel. The other extreme is a coil with many turns in which the conductor consists of a single strand. Figure 1 shows a typical form-wound coil in process after machine winding and after completion. This particular coil is a 5-turn coil with four strands per conductor. This coil is two strands wide. However, many coils are only one strand wide.

4. Coils having few turns are generally low-voltage high-current coils whereas coils having many turns are generally high-voltage low-current coils.

5. The coils of each phase may be connected one-circuit in which all coils are in series or they may be connected multi-circuit in various parallel series combinations.

6. The number of stator slots possible in a motor is determined by the stator diameter and allowable slot proportioning based

---

Paper 52-308, recommended by the AIEE Rotating Machinery and General Industry Applications Committees and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted April 25, 1952; made available for printing September 4, 1952.

C. E. Miller is with the General Electric Company, Schenectady, N. Y.

The author is grateful to D. S. Brereton for his assistance in making the economic study and preparation of data in Figures 8 and 9.

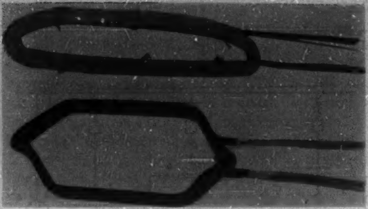


Figure 1 (left). Typical form-wound coil after machine winding and after completion. Five turns, four strands per turn

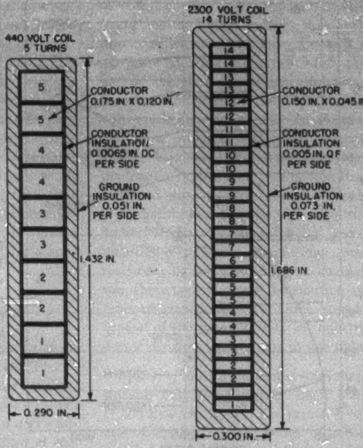

440 VOLT COIL
5 TURNS
5
5
4
4
3
3
2
2
1
1
CONDUCTOR 0.175 IN. X 0.120 IN.
CONDUCTOR INSULATION 0.0065 IN. DC PER SIDE
GROUND INSULATION 0.051 IN. PER SIDE
1.432 IN.
0.290 IN.
2300 VOLT COIL
14 TURNS
14
14
13
13
12
12
11
11
10
10
9
9
8
8
7
7
6
6
5
5
4
4
3
3
2
2
1
1
CONDUCTOR 0.150 IN. X 0.045 IN.
CONDUCTOR INSULATION 0.005 IN. QF PER SIDE
GROUND INSULATION 0.073 IN. PER SIDE
1.686 IN
0.300 IN.

Figure 2 (right). Cross sections of slot portions of form-wound coils for K6324, 60-hp 720-rpm motor

on good flux distribution. This ranges from one slot per phase per pole to as high as 15 slots per phase per pole. In general, motor designers prefer to use the same number of slots regardless of voltage within the range of practical voltage rating for the particular motor rating. However, in a majority of the borderline cases (voltage rating too high for hp rating) it becomes necessary to resort to a larger frame size in order to get the required number of slots necessary to accommodate the turns required for the higher voltage. In extreme cases the frame size may go up two or three sizes. As a compromise in the interest of keeping the number of turns down to a minimum the flux densities may be increased to the maximum the magnetic sections will stand. This increases the starting current resulting in higher mechanical stresses on the coils while at the same time the higher voltage coils are more flexible and less able to withstand these stresses.

Considering any particular hp rating as the voltage rating is increased, the limit is approached whereby the maximum number of turns of the smallest allowable conductor size per coil must be used and all coils of each phase must be connected in series (single circuit). This winding arrangement is the worst possible combination from the standpoint of mechanical strength. It is the winding arrangement that the designer is forced to resort to in those cases where too high a voltage rating is specified. A coil consisting of many turns of a single small size conductor means that the two conductors coming out of the coil for connection to other coils also will consist of the small conductor which is very vulnerable to vibrations encountered during starting and normal running.

Figure 2 compares cross sections of the slot portion of a 440-volt coil and a 2,300-volt coil for a type-*K* 60-hp 720-rpm motor. Note the relative size of the conductor strands, a ratio of over 3 to 1 in cross-sectional area. The high-voltage coil involves more turn-to-turn insulation and more ground insulation. The ratio of insulation to copper for the 2,300-volt coil is approximately 1.7 to 1 in contrast with 1 to 1 for the 440-volt coil. Thus, selection of too high a voltage rating results in a coil with a high ratio of insulation to copper. This results in a relatively softer, less rigid coil, which is harder to wedge securely in the stator slots, and harder to keep tight under conditions of continual heating and cooling, starting and stopping; it is more difficult to brace the coil ends, and the coil connections are more vulnerable to vibration. Contrast this with the cases where the proper voltage, 440 volts in this case, is selected. The coil conductor consists of much larger strands, over three times as large, and the coil connections are therefore larger; less over-all turn-to-turn and ground insulation is required, resulting in a harder coil which stays tight in the slot longer. As the limit in this direction, consider for a moment the extreme case of a single turn coil. Here the conductor would consist of many relatively large strands of copper; the coil connections would consist of this same multistrand conductor; there is a minimum of insulation on each strand, required only to prevent eddy current losses within the copper; therefore this coil is the hardest, most rigid, strongest coil mechanically that it is possible to build. The single turn coil is, of course, one extreme of the range of coil designs, and it is usually impractical to use on industrial motors. However, there are thousands of good coil designs in between this extreme case and the questionable small conductor multiple-turn coil so there is really no excuse for selecting a voltage rating which makes it necessary to use a compromise coil design.

Effect of Speed on Stator Windings

Figure 3 shows views of wound stators of a high speed 2-pole motor as compared with a lower speed 12-pole motor. In the case of the 2-pole machine the coil pitch span is very long, approximately 120 mechanical degrees compared to approximately 27 mechanical degrees for the 12-pole machine. The natural stiffness of the coils in the 12-pole machine is sufficient such that additional bracing is not required. The long coil ends of the 2-pole machine require a lot of bracing to support them adequately against vibration due to normal operation, high starting

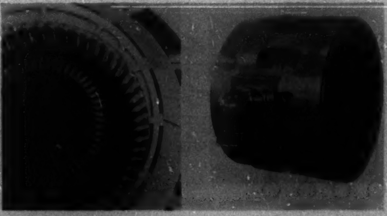


Figure 3 (right). Typical 2-pole stator winding and 12-pole stator winding. Note bracing of end turns on 2-pole machine necessary because of long coil span

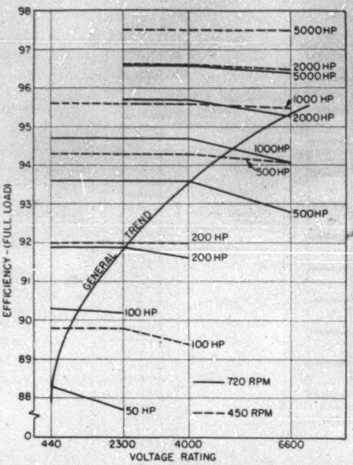


Figure 4 (left). Efficiency versus voltage rating for 0.8 power factor synchronous motors

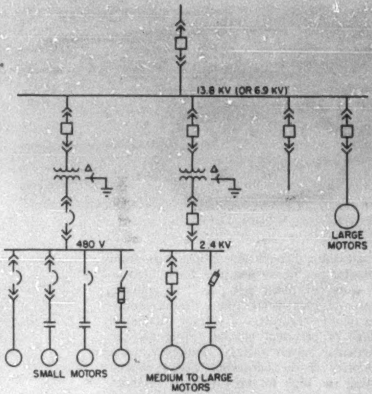


Figure 6 (right). Typical system with 13.8-kv primary voltage (case 2A 6.9 kv few special cases). Partial 1-line diagram

current, and fluctuating load currents. High-speed motors with fewer poles are inherently lower reactance motors than low-speed motors and consequently involve higher starting currents. Thus, there is a sort of double-barrelled effect in that the winding is inherently harder to brace properly and it permits a higher starting current, thereby subjecting the winding to more abuse. Thus it becomes evident that the selection of voltage may be influenced by the speed rating.

Synchronous Versus Induction

Synchronous motors have much larger air gaps than induction motors and therefore can be built with larger slots and fewer slots per pole without impairing the operating characteristic to the same degree as might occur on an induction motor. Rotor surface losses vary inversely as the air gap and the number of slots. The larger and fewer slots possible in synchronous motors result in better utilization of slot space—less slot space lost to insulation. The starting currents of synchronous motors are lower than for squirrel-cage induction motors and therefore the duty on the windings during starting is less for synchronous motors. Thus in many cases a particular voltage rating will be better suited for a synchronous motor than for an induction motor.

High-Potential Test

The standard high-potential test, by the American Standard Association rules, is twice rated voltage plus 1,000 volts. Thus motors rated 440 volts are given a high-potential test of 1,880 volts or 4.27 times rated voltage. At the other extreme a motor rated 13.2 kv is given a high-potential test of 27.4 kv which is only 2.07 times rated voltage. Of course, these two extreme voltage ratings should never be considered as possibilities for a particular hp rating. There will be, however, cases where 440 volts and 6,600 volts or 2,300 and 13,200 volts are possible choices. The point is that there is more insulation margin in a low-voltage motor than in a high-voltage motor. Therefore, all other things being equal, the low-voltage motor should be chosen.

Effect of Voltage on Efficiency and Power Factor (Induction Motors)

As the voltage rating is increased for a given hp and speed rating, a point is reached where it is necessary to resort to a compromise design, thereby making it impossible to obtain the same efficiency

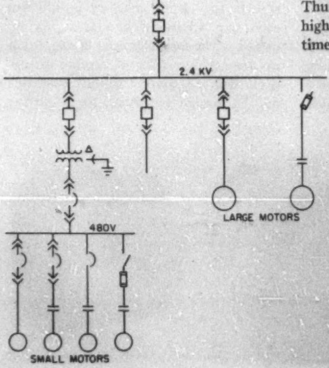
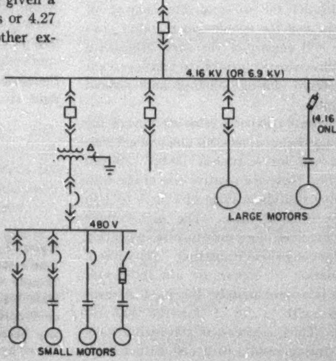


Figure 5 (left). Typical system with 2.4-kv primary voltage found in very small plants or where plant owner had no choice (purchased at 2.4 kv). Partial 1-line diagram

Figure 7 (right). Typical system with 4.16-kv primary voltage (case 3A 6.9 kv few special cases). Partial 1-line diagram



as can be obtained if designed for the proper voltage. This is quite evident in Figure 4 which shows the break in the efficiency curves as the rated voltage is increased for various hp ratings. The curve labeled "General Trend" is an attempt to draw a smooth curve through the breaks in the efficiency curve.

These efficiency-voltage curves are for synchronous motors. A review of similar data for induction motors will show the same general trend. Also a review of induction motor data reveals that power factor drops off very appreciably as the voltage rating is increased. Depending upon the public utility rate structure, power factor might warrant some consideration in selecting the voltage.

Application Considerations

The type of application should be given consideration in selecting the proper voltage rating especially in connection with borderline cases. Frequency of starting, type of load, whether steady or fluctuating, ambient conditions, and type of enclosure all warrant consideration. A higher rated voltage should never be selected, at the expense of winding mechanical strength, for motors that are to be subjected to more than the normal number of starts per day. The same is true to a degree for a highly fluctuating load. The high currents encountered during starting subject the windings to severe stresses, many times normal running stresses (forces on conductors vary approximately as the current squared), and the best possible winding is needed.

The ambient conditions and the type of enclosures are closely related, of course, inasmuch as the ambient surrounding the winding is dependent upon the enclosure. In some cases the ambient conditions may be the deciding factor in selecting the voltage. For example, a 300-hp 4,000-volt motor is entirely satisfactory in a clean dry ambient but it would be very undesirable to apply this motor where the ambient conditions are not ideal, for example, a splashproof motor outdoors or such applications as jordans and beaters in a paper mill where moisture and acid are prevalent. (This should not be misconstrued to mean that a lower voltage would be satisfactory in splashproof construction for these applications. The proper solution for such applications is, of course, adequate enclosures.) Enclosed motors of the type in which the windings are completely protected from the external atmosphere such as totally enclosed, enclosed fan cooled, or enclosed motors with surface coolers are much less vulnerable to winding failures than open motors. Thus there may be cases where the next higher voltage might be tolerated if an adequate enclosure is provided.

Applications involving part winding starting warrant special consideration in regard to voltage rating. In general, part winding starting is not feasible for relatively small high-voltage motors because of the necessity for multiple circuit windings, and consequently more turns per coils in most cases. The recommended minimum horsepower for a given voltage is therefore much higher than for a conventional motor.

Miscellaneous Considerations

In ratings where the choice is between 440-volts and high voltage, small 440-volt motors possess a great deal more utility, flexibility, and resaleability than do small high-voltage motors. It is to a user's advantage to be able to use a motor at any location throughout his plant. The resale value of a motor with the proper voltage rating for the hp rating is also better.

Motors of voltage ratings higher than indicated by good practice seldom are carried in stock. This may be of little importance except in an emergency. Then it can mean the difference between immediate delivery and several months' delivery.

Economic Considerations

The selection of the proper voltage rating for motors is extremely important if the minimum over-all installed cost is to be realized. Any cost comparison at different motor voltage ratings must take into consideration, of course, the cost of adequate starting equipment and that part of the plant power distribution system that is different for the two utilization voltages being considered. In the past it is quite probable that many economic studies were based on motor starters with inadequate interrupting ability (ten times full load motor current) and also old-style distribution systems (one large substation feeding the entire plant); consequently erroneous conclusions were reached. In the following cost comparison, three typical power distribution systems are considered. The three cases are typical of the vast majority of industrial power distribution systems in use today.

Case 1: Figure 5 shows a system that is suitable only for relatively small plants (7,500 kva and less). The primary voltage of 2.4 kv seldom would be recommended for a new plant. However, this case is included here because there are many existing plants with systems similar to this. In this case the cost comparison includes the motors, control, and substation for 440 volts, and motors and control for 2,300 volts.

Case 2: Figure 6 shows a system in which the primary voltage is 13.8 kv (or case 2A 6.9 kv in some few plants) and three utilization voltages, 440, 2,300 or 13,200, are available for the motors. This type of system is typical of large plants of 20,000 kva and up. The comparison of cost for the different utilization voltages includes everything back to the main

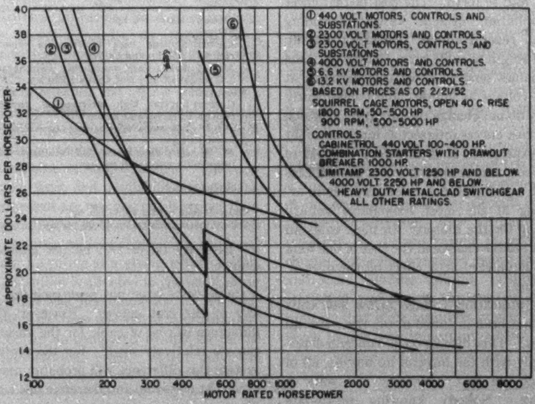


Figure 8. Approximate costs of motors, control, and substations as a function of motor hp

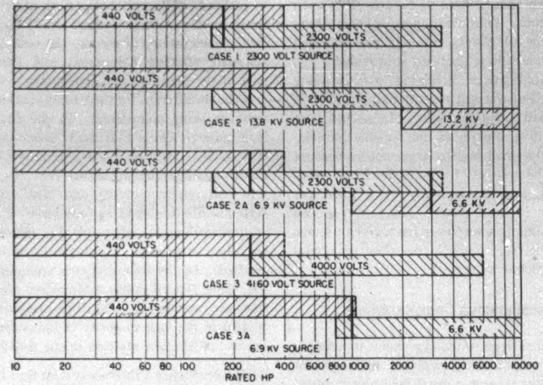


Figure 9. Recommended motor ratings—voltage versus hp

13.8-kv primary bus. Thus at either 440 or 2,300 volts the cost of motors, control, and substations is included.

Case 3: Figure 7 shows a typical system with 4.16-kv primary voltage (or case 3A 6.9 kv) and is commonly found in plants in the range of 5,000- to approximately 20,000-kva capacity. Other than the primary voltage this case is similar to case I.

There are many factors that affect cost, and any generalization in the form of cost curves must be well qualified. To name only a few of the variables: Motor costs vary with type and rated speed. Control costs vary as to type, full voltage, reduced voltage, reversing, dynamic braking, and so forth. The cost of the power distribution system varies depending upon general layout, size of substation, and length of cable runs. However, these variables tend to affect cost as much at one voltage as another and consequently they do not change the hp rating at which the cost curves intersect as much as might be expected. It is therefore possible to use cost curves applying to a general case with a practical degree of accuracy in arriving at the most economical voltage rating for the motors. In most cases an appreciable change in the cost of one item will have very little effect on the over-all picture.

A utilization voltage below 440 volts has not been included in the three typical cases. This discussion is concerned primarily with the integral hp motors, all of which are best served by 440 volts or above. Nobody knowing up-to-date power distribution practices would consider 220 volts in preference to 440 volts for general power. Small portable hand tools should be operated from a 110- or 220-volt circuit of course.

Figure 8 shows approximate cost of motors, control, and substations as a function of motor rated hp. These costs are based on the following:

1. Cost as of February 21, 1952, plus 10 per cent for estimating purposes.

2. Squirrel-cage motors, open 40 degrees centigrade, standard torques, 2-bearing, horizontal.

   1,800 rpm, 500 hp and below
   900 rpm, 500 hp and above

3. Control, full voltage starting, nonreversing.

   440 volt
      Cabinetrol* control center 400 hp and below
      Combination starters with drawout air circuit breakers above 400 hp
   2,300 volt
      Current-limiting high interrupting capacity control 1,250 hp and below
      MI 26 heavy-duty metal-clad switchgear—150 megavolt-amperes above 1,250 hp
   4,000 volt
      Current-limiting high interrupting capacity control 2,250 hp and below
      MI 26—250 megavolt-amperes above 2,250 hp
   6.6 and 13.2 kv
      MI 36 heavy-duty metal-clad switchgear—500 megavolt-amperes all ratings

4. Substations

   480-volt load center $9.00 per hp
   2.4-kv master substation $4.00 per hp

5. Diversity factor $1^{1}/_{3}$ hp per kilovolt-ampere.

Cable costs have not been included in any of the cost curves. However, the difference in cost of cable for the various cases is relatively small. For the typical system this difference will probably never exceed $1.00 per hp and therefore has very little effect on the point of intersection of the curves.

Many important conclusions can be drawn from the cost curves in addition to the points of intersection of the various curves. The relative slope of the curves is perhaps of as much or more significance than the point of intersection. For example, curves *1* and *2* intersect at 175 hp, which indicates that at this hp 440-volt motors, control, and substation cost the same as 2,300-volt motors and control. However, these curves also say that at 100 hp a 2,300-volt motor and control ($45.00 per hp) cost 32 per cent more than a 440-volt motor, control, and substation. Thus the purchaser who insists on buying a 100-hp 2,300-volt motor is not only getting a motor with less rugged stator windings but is paying a premium for this undesirable choice. Again, considering curves *1* and *2*, a 400-hp 440-volt motor installation costs 40 per cent more than a 2,300-volt motor installation. If curves *1* and *3* are considered (case II which involves substations for both 440 volts and 2,300 volts), then at 100 hp the 2,300-volt motor installation ($49.00 per hp) costs 44 per cent more than the 440-volt motor installation. Likewise a 400-hp 440-volt motor installation costs 16.5 per cent more than a 2,300-volt motor installation.

Attention is directed to the sharp break in the curves at 500 hp. This is due to the difference in cost of 1,800-rpm and 900-rpm motors. The two speeds were used in preparing the curves because 1,800-rpm motors are used more commonly in the smaller hp ratings, whereas 900-rpm motors are more common in the larger ratings.

One interesting observation in connection with cost curves *1*, *2*, *3*, and *4* is that the type of control selected for this general case makes the intersection point fall at a lower hp rating than for most any other type of control. For example, if reduced voltage starters are considered, the increase in cost of 2,300-volt control is much more than for 440-volt control. Consequently, where reduced voltage control is involved, cost favors the use of 440-volt motors at higher hp ratings than are indicated in Figure 8 for full voltage starting equipment. This same thing holds true for control involving such features as reversing and dynamic braking.

Curve *1* in Figure 8 for 440-volt equipment covers ratings up to approximately 1,000 hp. This should not be misconstrued to mean that 440-volt equipments in these larger hp ratings are recommended ordinarily. Curve *1* is extended to include the intersection point with curve *5* because of those installations where only 440 volts and 6,600 volts are available.

---

* Registered trade-mark of General Electric Company.

In general, 440-volt equipments above 400 hp require special consideration from the standpoint of available starting equipment.

Curves *2* and *3* are terminated at 3,500 hp on the basis that this is about the largest motor compatible with available 2,300-volt switchgear starters and with the rating of substation required to serve a motor of this rating satisfactorily. On the same basis, curve *4* for 4,000-volt motors is terminated at 5,000 hp. However, the maximum hp rating of motor that can be served satisfactorily by a given size substation will vary of course, depending upon whether full voltage or reduced voltage starting is employed and whether a synchronous, squirrel-cage, or wound-rotor induction motor is used. This cost study was not carried beyond 5,000 hp. Motors above 5,000 hp are best served at 13.2 kv, 11 kv, or 6.6 kv.

Voltage Versus Horsepower Chart

To summarize, the various factors that should be considered in selecting the proper motor voltage are:

1. Mechanical strength of stator winding.
2. Effect of speed on stator winding.
3. Synchronous versus induction.
4. High-potential test.
5. Efficiency and power factor.
6. Application considerations.
7. Economic considerations.

In an effort to combine all of these factors into the most convenient form for ready reference, a bar chart, Figure 9, has been prepared. In using this chart, reference should be made to the typical power distribution 1-line diagrams previously shown.

The heavy vertical lines that extend across two voltage bars indicate the hp at which the costs are the same at either voltage as derived from Figure 8. Note the absence of the vertical line in case II for 2,300 volts and 13.2 kv. Reference to Figure 8 will show these two curves do not intersect.

The minimum hp indicated in each voltage class in Figure 9 is based on the recommended minimum rating that is consistent with a good motor design. Note that these hp ratings are the recommended minimum under the most favorable conditions as to type of motor, rated speed, or application. As discussed previously, the recommended minimum hp rating for a given voltage rating is affected by the type of motor, whether synchronous or induction, by the rated speed, and by application considerations. Accordingly there are many specific cases where these lower limits should be adjusted in the upward direction. For example, high-speed 2-pole induction motors definitely would fall in this category and should be adjusted upward. It should be borne in mind that these recommended minimum hp ratings are not peculiar to any one motor manufacturer; the same is true of all motor manufacturers.

Table I

Voltage Rating		Minimum HP
440 volts		3/4 to 1
2,300 volts		150 to 200
4,000 volts		250 to 300
6,600 volts		800 to 1,000
11 or 13.2 kv	Syn	2,000 to 2,500
	Induction 2-pole	not recommended
	4- and 6-pole	5,000
	all others	3,000 to 3,500

As a further aid in determining these lower hp limits for the various conditions as to type of motor, speed, and application, these limits may be expressed to cover the range corresponding to the most favorable conditions to the most unfavorable conditions. The most favorable conditions would be a slow-speed synchronous motor operating under ideal conditions and at the other extreme would be a high-speed 2-pole induction motor operating under conditions similar to those normally encountered in a power station. Table I gives these limits.

In Figure 9 the dotted lines indicating maximum hp ratings for a given voltage are based primarily on considerations other than motor design. So far in this paper the emphasis has been placed on the evils of selecting too high a voltage for a given hp rating. It is, of course, possible to err in the other direction but not nearly so probable, especially from the standpoint of motor design. The economic limit and availability of starting equipment are reached much sooner than the motor design limit and should be used as the main guide in determining the maximum hp rating at a given voltage rating.

Attention is directed to the fact that in case 3A the maximum hp for 440-volt motors has been selected at approximately 1,000 hp as compared to 400 hp in the other 440-volt cases. When only 440 volts and 6,600 volts are available as utilization voltages of necessity, a compromise must be made within the range of approximately 500 to 1,000 hp. In this case the economic break-even point is approximately 1,000 hp and therefore the 440-volt bar has been extended to include this point.

All of this paper has been based on the assumption that when making a choice between two voltages, either voltage is available near the motor location. While this probably will be true in a majority of cases, especially in new installations, there will be many cases where even though both voltages are available, available capacity at each voltage may have a bearing on the selection. If capacity is available only at a voltage which is considered too high for a good motor design (borderline case), in extreme cases it may be advantageous to consider the cost of the higher voltage motor as compared to the cost of a low-voltage motor and an individual transformer. In this case the motor and transformer would be switched as a unit, the starter being connected on the high-voltage side of the transformer. Also, in cases like this, where system capacity is available at only one voltage, the possibility of future system changes or modernization programs should not be overlooked whereby the proper voltage would be made available.

It will be noted that in the cases shown in Figure 9 there is no case that makes a direct comparison of 2,300 volts and 4,000 volts. That is a broad subject within itself. However, one point that is worth mentioning here is that within the range of approximately 1,250 to 2,500 hp, 4,000 volts has the advantage of extending the range of current-limiting at high interrupting capacity control. This is an important consideration in the case of applications involving frequent starting or control with additional functions such as reversing, reduced voltage starting, and dynamic braking.

In conclusion, there is no excuse for selecting and applying motors of too high a voltage rating for a given horsepower rating, especially when over-all costs favor the proper selection. Also, an improper selection of voltage rating is inviting shorter motor life, increased maintenance, and more winding failures. However statistical records over the last few years indicate that more attention now is being given to this very important subject, the proper selection of motor voltage.

No Discussion

Stabilization of a Servomechanism Subject to Large Amplitude Oscillation

EMILE S. SHERRARD
ASSOCIATE MEMBER AIEE

Synopsis: The saturation of an amplifier is found to cause large amplitude oscillations in a conditionally stable servomechanism. The employment of a suitable nonlinear element in the stabilizing network may prevent such oscillations. For a particular system containing a saturable amplifier, frequency response analyses are used to find the magnitude and frequency of the oscillation and to determine the design of a nonlinear filter which will prevent large amplitude oscillation. Laplace transformation theory is employed to justify the use of a frequency response analysis, and to prove that the combination of a saturable amplifier and an unstable inner loop will cause oscillations in both conditionally and unconditionally stable systems. Results obtained from an analogue computer are compared with the results of the frequency response analyses.

THE analysis and synthesis of servomechanisms frequently is facilitated by assuming that the behavior of the physical system is the same as the behavior of an ideal system identical with the physical system except that it contains no saturable amplifiers, backlash, or other "nonlinearities." For some systems there is little difference in the behavior of the physical and ideal systems. For other systems, the physical and ideal systems may differ appreciably in performance, or the physical system may be unstable while the ideal system is stable.

One particular type of system for which the physical system is frequently unstable and the ideal system stable is the "Nyquist stable" or "conditionally stable" system. Such a system is characterized by the fact that the ideal system becomes unstable if the system gain is sufficiently reduced from its nominal value. The physical system is usually stable for small signal operation. However, it frequently becomes unstable and exhibits large amplitude oscillations when one or more amplifier tubes saturate as a result of large signals or transient phenomena that occur when the system is put into operation.

The likelihood of oscillation in conditionally stable systems is so great that such systems are used infrequently for feedback amplifiers.[1] In the field of servomechanisms, conditionally stable systems often result from the requirements of high gain and simple or economical stabilizing networks.

In this paper, a frequency response analysis will be used to determine the stability of a physical system containing a saturable amplifier. The same method of analysis will be used to determine the design of a nonlinear stabilizing filter which will prevent large amplitude oscillations in the system. The frequency response analysis has been explained thoroughly in a previous paper.[2] As employed in the present instance, this method of analysis assumes that the steady-state or asymptotic oscillation of the system is periodic and that the fundamental frequency term is the only important term of the Fourier series representing the periodic oscillation.

A frequency response analysis resembles a graphical analysis in that its results are given in numerical rather than literal form. The numerical values of parameters in a particular system must be assigned before the analysis may be made. Therefore, a particular system was chosen for analysis and numerical values assigned its parameters. Much of the remainder of this paper will be concerned with this particular system. However, the method of analysis is independent of the system analyzed. If applied to other conditionally stable systems, the same method may determine the characteristics of system oscillation and may determine the design of nonlinear stabilizing networks which will prevent large amplitude oscillation.

Description of Physical System

The block diagram of the system assumed for analysis is shown in Figure 1. The system consists of a cascade connection of a motor, a power amplifier driving the motor, a lag type of stabilizing network, a voltage amplifier, and a differential which subtracts $c(t)$, the output of the motor, from $r(t)$, the reference input, and supplies the result, $e(t)$, to the voltage amplifier.

The system is assumed to contain no nonlinearities other than the saturable voltage amplifier. The transfer function of each element is written below that element in Figure 1. The over-all transfer function is the product of the individual transfer functions and it also is given in Figure 1. This transfer function is written in both literal form and in the numerical form employed in the frequency response analyses. In this paper, the term "transfer function" denotes the ratio of the Laplace transforms of two quantities when the initial conditions appearing in each transform are equal to zero.

The nomenclature of Figure 1 is that proposed by the AIEE Subcommittee on Terminology and Nomenclature of the Feedback Control Systems Committee.[3] In particular, the reader should note that a quantity which is a function of time is represented by a lower-case letter while the Laplace transform of the quantity is represented by the capital letter corresponding to the lower-case letter. For example, c or $c(t)$ is used to denote the motor output, while C or $C(s)$ is used to denote the Laplace transform of the motor output.

The saturation characteristics of the voltage amplifier are given in Figure 2. The magnitude of the amplifier output is A times the magnitude of the amplifier input for input magnitudes less than e_s/A. For larger magnitudes of input the output magnitude is e_s.

Frequency Response Analysis of System with Saturable Amplifier

This section will be concerned with the determination of the approximate amplitude and frequency of the self-sustained oscillations of the physical system of Figure 1.

From the results of Appendix I, the poles of $C(s)$, the Laplace transform of the motor output, are of the first order and lie on the imaginary axis or within the left half of the complex plane. If the properties of $C(s)$ are such that the ordinary inversion formula for $C(s)$ holds, then $c(t)$, the motor output, may be written as a finite or infinite series of exponentials.[4] The arguments of these exponentials will be the product of time and the values of s at which $C(s)$ has poles. As time in-

---

Paper 52-277, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted February 11, 1952; made available for printing August 12, 1952.

EMILE S. SHERRARD is with the Naval Research Laboratory, Washington, D. C.

The author wishes to acknowledge the valuable suggestions and guidance of Dr. A. H. Fox of Union College. Most of the material contained in this paper is taken from a master's thesis prepared under his supervision. Dr. S. J. Jennings and C. H. Thomas of the General Electric Company aided the preparation of this paper by their constructive comments, and J. E. Gerngross of the General Electric Company supervised the analogue computer study.

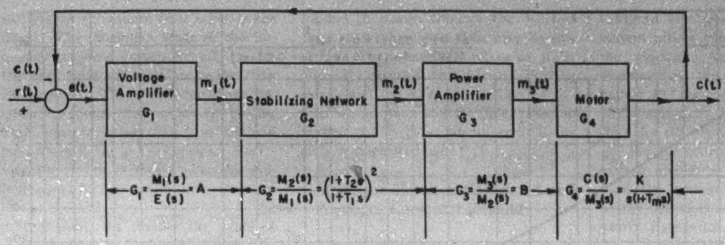


Figure 1. Block diagram for typical servo system

$$\frac{C(s)}{E(s)}=\frac{ABK\,(1+T_2 s)^2}{s(1+T_1 s)^2\,(1+sT_m)}=\frac{930(1+0.5s)^2}{s(1+5s)^2(1+0.03s)}=ABK\frac{N(s)}{D(s)}$$

creases the only important terms in the series for $c(t)$ will be the constant and sinusoidal terms corresponding to poles of $C(s)$ on the imaginary axis. Therefore, the steady-state or asymptotic response of the motor output may consist of a constant plus sinusoidal terms.

If $c(t)$ were a single sinusoid, the characteristics of the self-sustained oscillations could be determined readily. A consideration of the harmonic generating property of the saturable amplifier shows that a single sinusoid cannot be the exact form of $c(t)$. However, for a system such as this, a single sinusoid may be a good approximation to the actual form of $c(t)$. To understand how this may happen, suppose that $c(t)$ is a simple sinusoid. Even if the amplitude of $c(t)$ is so large that the output of the saturable amplifier is very nearly a square wave, the fundamental of this output will be larger in magnitude than any of the harmonics. The system elements between the amplifier output and the motor output constitute a low-pass filter. The fundamental component of the amplifier output is attenuated less by this filter than are the harmonic terms of this output. If this low-pass filtering effect is sufficiently large, the harmonic components of $c(t)$ may be considered to be negligible. In the remainder of this section, $c(t)$ will be assumed to be a simple sinusoid. The accuracy of this assumption is investigated in a later section.

Since the ideal system is stable, any self-sustained oscillation must be sufficiently large to cause saturation of the voltage amplifier. If $c(t)$ is a sinusoid, the voltage amplifier input will be $-c(t)$ and the output of this amplifier will have a fundamental component given by $-ASc(t)$. Here A is the gain of the unsaturated amplifier and S is a factor ≤ 1 which accounts for the saturated operation of the amplifier. It has been assumed that the harmonic components of the amplifier output may be neglected in the determination of the motor output. The transfer function between the amplifier output and the motor output is, from Figure 1, $KBN(s)/D(s)$. In the usual complex notation applicable to sinusoidal quantities the relation between the transforms of the fundamental of $m_1(t)$ and $c(t)$ may be written as

$$C(j\omega)=\frac{KBN(j\omega)}{D(j\omega)}M_{1f}(j\omega) \qquad (1)$$

where $M_{1f}(j\omega)$ denotes the transform of $m_{1f}(t)$, the fundamental of $m_1(t)$.

Substituting $M_{1f}(j\omega)=-ASC(j\omega)$ we have

$$C(j\omega)=\frac{-KASBN(j\omega)}{D(j\omega)}C(j\omega) \qquad (2)$$

A nontrivial solution of equation 2 requires

$$\frac{KASBN(j\omega)}{D(j\omega)}=-1 \qquad (3)$$

To meet this condition, the frequency of $c(t)$ must be such that the angle of $N(j\omega)/D(j\omega)$ is 180 degrees. Also the amplitude of $c(t)$ must be such that the absolute value of the left member of equation 3 is unity.

Figures 3 and 4 are plots of the absolute magnitude and angle of $KABN(j\omega)/D(j\omega)$ versus ω. An examination of Figure 4 shows that ω must be either 0.26 or 1.65 radians per second if the angle is to be 180 degrees. The value of S as a function of amplitude is determined in Appendix II and plotted in Figure 5. This plot is of S versus the per-unit amplitude of $e(t)$. The per-unit amplitude of $e(t)$ is defined as the amplitude of $e(t)$ divided by e_s/A. For $\omega=0.26$, if the absolute value of $KASBN(j\omega)/D(j\omega)$ is to be 1, S must be 8.91×10^{-4}. Then $c(t)$ must have an amplitude of 1,440 per unit. Similarly the other solution of equation 3 is found to require that $c(t)$ be 15 per unit in amplitude and of 1.65 radians-per-second frequency.

Polar plots of $KASBN(j\omega)/D(j\omega)$ are given in Figure 6 for values of S corresponding to ≤ 1, 15, and 1,440 per-unit amplitudes of $c(t)$.

Of the two possible solutions of equation 3 one probably will be a stable oscillation, and the other an unstable oscillation. The oscillation of 1,440 per unit and 0.26 radian per second may be the stable oscillation while the 15 per-unit, 1.65 radians-per-second oscillation may be unstable. The determination of the stability of the possible oscillations is beyond the scope of this paper. This section is primarily concerned with the approximate determination of the possible oscillations of the system.

Nonlinear Filter for Stabilizing Nonlinear System

The linear stabilizing filter is entirely adequate for a linear or ideal system but it fails to stabilize the physical system against large amplitude oscillations. Stabilization in this case may be obtained by a nonlinear filter. The filter considered here has a clamping or limiting circuit in

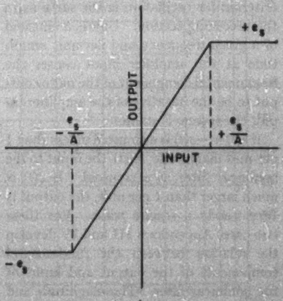


Figure 2. Saturation characteristic of voltage amplifier

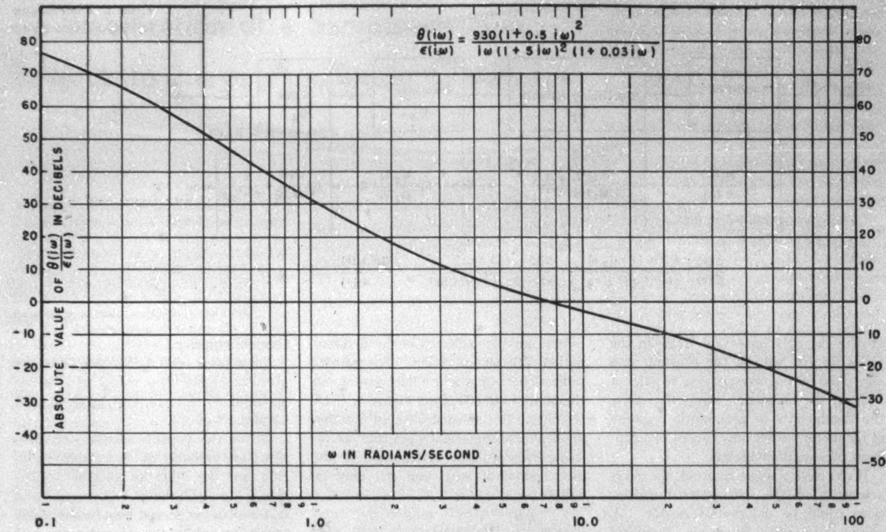


Figure 3. Amplitude of transfer function of servo versus frequency

the first section of the 2-section stabilizing network. The clamping circuit prevents the capacitor voltage in this section from exceeding a certain positive or negative value. Appendix III contains a description of the detailed operation of the filter.

The analysis of Appendix I when adapted to this case shows that $C(s)$, the transform of the motor output, contains no poles in the right half plane and may have no more than single-order poles on the imaginary axis. Therefore, the following frequency response analysis assumes that for any self-sustained oscillation the form of $c(t)$ is a sinusoid. The criterion for oscillation is the same as in the preceding section. Unless a sinusoid of a given frequency and per-unit amplitude at the amplifier input causes the fundamental component of the motor output to be the negative of the amplifier input, the system is stable.

If $e(t) = -c(t)$ is a sinusoid of less than 1 per-unit magnitude, $m_1(t)$, the input to the nonlinear filter, is a sinusoid. If $c(t)$ is much larger than 1 per unit, the output is very nearly a square wave. For these two cases Appendixes III and IV develop the relation between the fundamental components of the output and input of the nonlinear filter. The magnitude and phase characteristics of the nonlinear filter are shown by the curves of Figures 7, 8, 9, and 10. The parameter μ of these curves is defined as the ratio E_s/e where E_s is the "clamping voltage" of the capacitor and e is the peak value of the filter input voltage. If the amplitude of amplifier input is larger than 1 per unit, e is equal to e_s, the maximum output of the saturated amplifier.

For a sinusoidal $e(t)$, Figures 11 and 12 show for various frequencies in polar form the ratio of the amplitudes of the fundamental of $c(t)$ and $e(t)$ and the phase angle between the fundamental of $c(t)$ and $e(t)$. Figure 11 applies for $e(t)$ less than 1 per-unit amplitude (sine wave input to filter) and Figure 12 applies to a large per-unit amplitude of $e(t)$ (square wave input to filter). The curves of Figures 11 and 12 may be used to determine for what values of E_s/e_s the system is stable. As an example, consider the case in which E_s/e_s is taken equal to 0.1. For amplitudes of $e(t)$ less than 0.1 per unit, the peak output of the voltage amplifier is less than 0.1 e_s and is less than E_s, the clamping voltage of the capacitor. Then the clamping circuit is inactive, the voltage amplifier is unsaturated, and the frequency response is given by the $\mu \geq 1$ curve (linear system characteristic) of either Figure 11 or 12. As $e(t)$ increases further to 1 per unit, the clamping circuit operates but the voltage amplifier remains unsaturated. The frequency response curve for $e(t)$ greater than 0.1 per unit and less than 1 per unit will lie between the $\mu \geq 1$ and the $\mu = 0.1$ curve of Figure 11. For $e(t)$ greater than 1 per unit, the input to the filter is no longer sinusoidal. For $e(t)$ greater than 2.5 per unit, the filter input voltage is nearly a square wave and the frequency response curve is very nearly S times the $\mu = 0.1$ curve of Figure 12. For $e(t)$ between 1 per unit and 2.5 per unit, the phase angle of the frequency response curve probably will be between the angles given by the $\mu = 0.1$ curves of Figures 11 and 12. If this is the case, none of the frequency response curves corresponding to the various per-unit amplitudes of $e(t)$ will pass through the $-1+j0$ point. Therefore, the system is stable.

Equation 24 may be used to calculate the approximate value of E_s/e_s necessary for stabilizing the system. In using this equation μ is replaced by E_s/e_s and ω_1 by ω_2.. ω_2 is taken as the value of ω at which the phase margin of the linear system is 0 and increases as ω increases. This choice of E_s/e_s causes the frequency response curves for $e(t)$ amplitudes between E_s/e_s and 1 per unit to lie between the $\mu \geq 1$ and $\mu = E_s/e_s$ curves of Figure 11. For $e(t)$ greater than 1 per-unit saturation of the voltage amplifier will cause the angle of the frequency response curves to lie between the values given by the $\mu = E_s/e_s$ curves of Figure 11 (sine wave input to the filter) and Figure 12 (square wave input to the filter). The maximum angle of the frequency response curves for $e(t)$ greater than 1 per unit will therefore be less than 180 degrees. Therefore, none of the frequency response curves for $e(t)$ greater than, equal to, or less than 1 per unit will pass through the $-1+j0$ point. Therefore, the system is stable. For this system, $\omega_2=1.6$, $T_1=5$, and E_s/e_s is calculated to be 0.124.

In general, the smaller the value of E_s/e_s, the less the likelihood of large amplitude oscillations. The value of E_s/e_s selected depends on the co-ordination of degree of stability, amount of overshoot permitted for a step function input to the system, the range of linear response required, and any other important design considerations. An upper value of E_s/e_s is determined by the intersection of $\mu \geq 1$ and the $\mu=E_s/e_s$ curve of Figure 12. The value of E_s/e_s must be such that the $\mu=E_s/e_s$ curve of Figure 12 does not join the $\mu \geq 1$ curve at a point where the phase angle is 180 degrees or greater.

If E_s/e_s is greater than this value, the intersection of the $\mu=E_s/e_s$ and $\mu \geq 1$ curves will have a phase angle greater than 180 degrees for a square wave input to the filter. Then a value of S may be found such that the $\mu=E_s/e_s$ curve of Figure 12 passes through the $-1+j0$ point and the system then may be unstable. The intersection of the $\mu=E_s/e_s$ and $\mu \geq 1$ curves occurs at $\omega=\omega_l$. Putting $\omega_l=\omega_2$ and $\mu=E_s/e_s$ in equation 14 determines an upper limit for E_s/e_s. For the system of this paper an upper limit for E_s/e_s is calculated to be 0.19.

Comparison Between Calculated and Analogue Computer Results

The system of Figure 1 was set up on analogue computer with $A=500$ and $BK=1.86$. Figure 13 shows the response to a large step input of the system containing the saturable voltage amplifier. The last part of the figure shows the steady-state wave shapes obtained after the step function was removed. Measured values of the output oscillation are 1,880 per-unit amplitude and 0.29 radian per second. Calculated values are 1,440 per unit and 0.26 radian per second. The asymmetrical half waves were obtained in Figure 13 for $m_1(t)$, the output of the saturable amplifier, and $e_c(t)$, the voltage across the capacitor of the first section of the stabilizing network. This asymmetry was caused by an unbalance in the diode limiter circuit which was used to limit $m_1(t)$.

Figure 14 shows the response of the system with a clamping circuit limiting the capacitor voltage. The largest overshoot in this instance is approximately 130 per cent of the final position of the output. In this figure E_s/e_s has the value 0.2.

When the system was in a state of self-excited oscillation, E_s/e_s was decreased until the system became stable. The border line of stability was found to be nearly $E_s/e_s=0.2$. For $E_s/e_s=0.3$ and 0.5, $c(t)$ was a sinusoid of 0.68 and 0.39 radian per second respectively.

Conclusions

1. Large amplitude oscillations of many conditionally stable systems may be prevented by the use of a nonlinear filter for stabilization.

2. The important parameter in the design of the nonlinear filter is the value of E_s/e_s. This ratio should be smaller than the value of μ given by the solution of equation 14 for $\omega_l=\omega_2$. ω_2 is the value of ω at which the phase margin of the linear system is zero and increasing as ω increases. A possible rule of thumb is to take E_s/e_s equal to μ as determined from equation 24 for $\omega_l=\omega_2$.

3. The results of the analogue computer study of the system justifies the assumption that the motor output is nearly a sinusoid when the system describes self-excited oscillations.

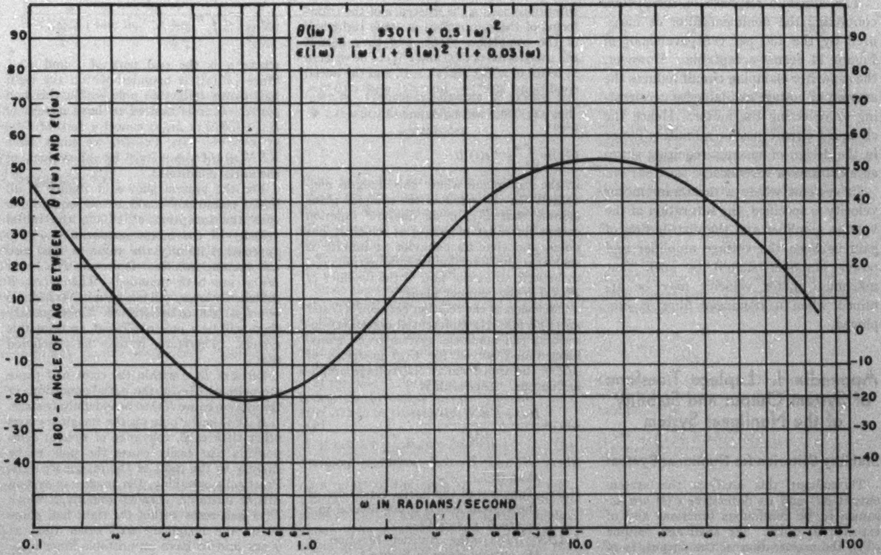


Figure 4. Phase angle of transfer function of servo versus frequency

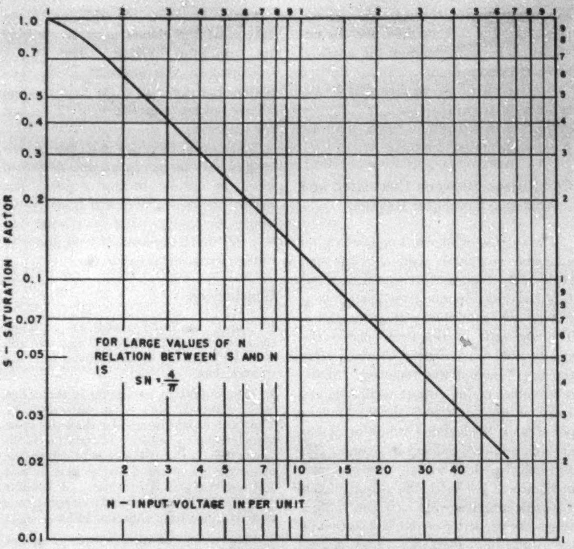


Figure 5. Saturation characteristics of voltage amplifier

Discussion

The dynamic response of the system containing the nonlinear filter as measured by the 130 per cent overshoot of Figure 14 seems acceptable. However, the capacitor clamping circuit reduces the amount of voltage available for accelerating or reversing the motor. Hence the clamping circuit may be a limiting factor in the design of systems requiring sharp accelerations or reversals.

In systems where a maximum motor velocity is specified, the saturation of the voltage amplifier and the distribution of gain between the voltage amplifier and power amplifier should be such that maximum motor velocity may be attained when a nonlinear filter is employed.

Appendix I. Laplace Transform of System Output and Stability of the Nonlinear System

Stability Criterion for Nonlinear System

Throughout this analysis the system output $c(t)$ and its derivative $c'(t)$ are assumed to be continuous functions and of exponential order. We shall also assume $c''(t)$, the acceleration of the output, is of exponential order and at most point-wise discontinuous. The exponential order assumption means that the absolute value of these functions is less than $M\epsilon^{\alpha t}$ where M and α are positive constants. These assumptions assure the existence of the transforms of these functions in some right half of the complex s plane, and they permit the transforms of $c'(t)$ and $c''(t)$ to be written in terms of $sC(s)$, $s^2C(s)$, s, s^2, and the initial conditions $c(0^+)=\lim_{t\to 0^+} c(t)$ and $c'(0^+)=\lim_{t\to 0^+} c'(t)$ $C(s)$ will be defined by

$$C(s)=\int_0^\infty \epsilon^{-st}c(t)\,dt$$

in the half plane where this integral converges. In the whole of the complex plane assume that $C(s)$ is an analytic function except for a set of points at which it has poles, and that its behavior at infinity is such that the inversion integral permits $c(t)$ to be written as the sum of the residues of $\epsilon^{st}C(s)$ in the complex plane.[4]

Exclusion of the relation between $e(t)$ and $m_1(t)$ permits the differential equations describing the nonlinear system to be transformed and solved for $C(s)$ in terms of $M_1(s)$, the transform of the voltage amplifier output. The result is

$$C(s)=\frac{I(s)+KBN(s)M_1(s)}{D(s)} \tag{4}$$

where K and B = real constants greater than 0

$I(s)$ = a function of the initial conditions of the system and is a polynomial in s as are $N(s)$ and $D(s)$

$KABN(s)/D(s)$ = the transfer function of the system given in Figure 1

The system considered here is the zero position error type and therefore $KABN(s)/D(s)$ has a single order pole at the origin.[5]

If the system is to be acceptable, its output $c(t)$ must, for $r(t)\equiv 0$, approach zero as time t approaches infinity. It must exhibit this behavior for all initial conditions. If it behaves in such fashion it is called a stable system. If it does not behave in such a fashion it is said that the unstable system is in a state of self-sustained oscillation or that it describes self-sustained oscillations. The criterion for stability of the nonlinear system is whether or not the poles of $C(s)$ as given by equation 4 lie on the imaginary axis or within the right half plane for $r(t)\equiv 0$. The poles of $C(s)$ for the nonlinear system are the poles of $M_1(s)$ and the zeros of $D(s)$, provided a zero of $N(s)$ does not cancel a pole of $M_1(s)$ or a zero of $I(s)+KBN(s)$ $M_1(s)$ does not cancel a zero of $D(s)$. The zeros of $D(s)$ and $N(s)$ are independent of the initial conditions in the system. They may be expected to vary with the tolerances of elements in the system. On the other hand $M_1(s)$ and $I(s)+KBN(s)$ $M_1(s)$ are functions both of the initial conditions and tolerances of the system. Therefore, one expects that for all possible initial conditions $I(s)+KBN(s)$ $M_1(s)$ will not have zeros that cancel the zeros of $D(s)$ on the imaginary axis or in the right half plane.

Effect of Unstable Inner Loop on System Stability

It can be shown that within the right half plane $I(s)+KBN(s)$ $M_1(s)$ cannot have a zero for all possible initial conditions. Note first that, because of the saturation of the voltage amplifier, $|m_1(t)|\leq e_s$. Then

$$|M_1(s)|\leq\int_0^\infty |m_1(t)\epsilon^{-st}|dt \text{ and } |M_1(s)|\leq\frac{e_s}{x}$$

where x is the real part of s and $x>0$. Since $|M_1(s)|$ is bounded within the right half plane, it has no pole within this half plane. If it is desired to have a zero of $I(s)+KBN(s)$ $M_1(s)$ cancel a zero of $D(s)$ at $s=s_0=x_0+iy_0$, $(x_0>0)$, we must have $KBN(s_0)M_1(s_0)=-I(s_0)$ for all values of the initial conditions.

For the system shown in Figure 1, all initial conditions could be set equal to 0 with the exception of $c'(0^+)$, the initial velocity of the motor output. As $c'(0^+)$ approaches infinity the value of $I(s_0)$ also will become infinite. However, $N(s_0)$ and $M_1(s_0)$ are both bounded. Then $I(s_0)$, if sufficiently large, will not permit $N(s_0)M_1(s_0)=-I(s_0)/BK$ to be satisfied. Consequently there will be a pole of $C(s)$ at $s=s_0$ for this system. Therefore, it may be concluded that:

Zeros of $D(s)$ within the right half plane are poles of $C(s)$, if the initial conditions in the system cause $I(s)$ to be indefinitely large.

If $D(s)$ has a pole on the imaginary axis other than at 0, tolerance of system components can easily cause the pole to be slightly to the right of the imaginary axis. Practically speaking, the nonlinear system will be unstable. Linear systems in which $D(s)$ has zeros within the right half plane or on the imaginary axis other than at 0 are said to have an unstable inner loop. Thus, the nonlinear system may be considered to be unstable if the linear system

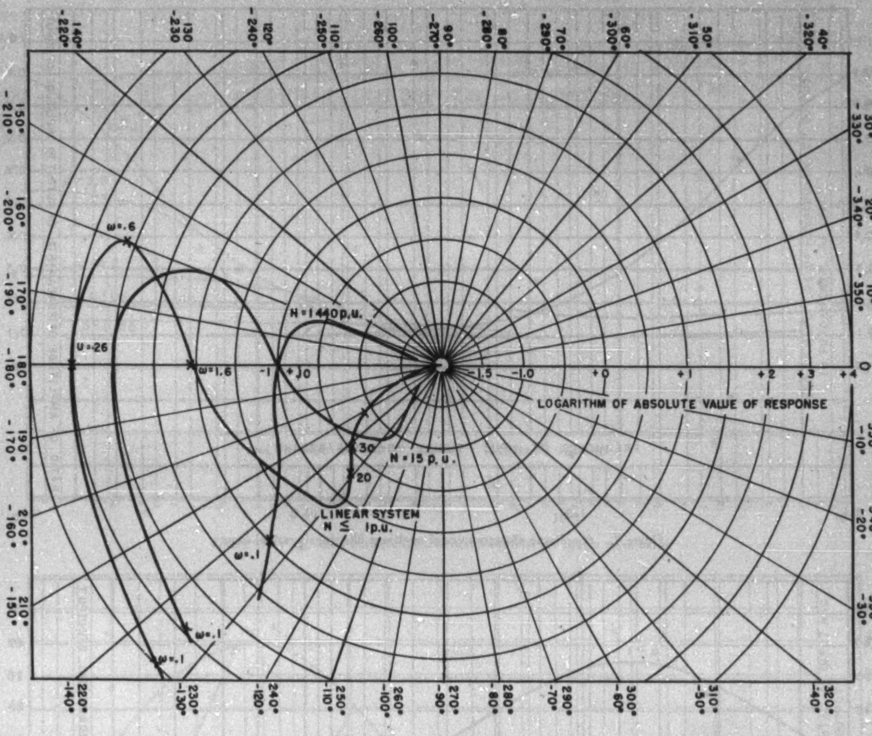


Figure 6. Frequency response of system for 1.0 per-unit, 15 per-unit, and 1,440 per-unit amplitudes

has an unstable loop. Hereafter any inner loops of the linear system will be considered stable.

Order of Poles of C(s) Lying on the Imaginary Axis

It has been shown that $M_1(s)$ has no poles within the right half plane. It may have poles on the imaginary axis. However, these poles will not be of an order greater than 1. To demonstrate this last statement observe that $|M_1(s)| \leq e_s/x$ where $x>0$ is the real part of s. If $M_1(s)$ had a pole of order greater than 1 on the imaginary axis at $s_0=0+iy_0$, then $\lim_{s\to s_0}|(s-s_0)\,M_1(s)| \to \infty$ as $s\to s_0$ in any manner whatsoever. However

$$|(s-s_0)M_1(s)| \leq \frac{e_s}{x}\sqrt{(x-x_0)^2+(y-y_0)^2}$$

$$=\frac{e_s|x-x_0|}{x}\sqrt{1+\left(\frac{y-y_0}{x-x_0}\right)^2}$$

If we let $s\to s_0$ along a straight line of slope k lying in the right half plane and terminating on the imaginary axis, then

$$\frac{y-y_0}{x-x_0}=k, \text{ and } \lim_{s\to s_0}|(s-s_0)M_1(s)|$$

$$\leq e_s\sqrt{1+k^2}$$

Thus $M_1(s)$ has no pole or order >1 anywhere on the imaginary axis. Assuming that $D(s)$ has no zeros on the imaginary axis other than at the origin, it may be concluded that:

Except at the origin, the poles of $C(s)$ on the imaginary axis are of the first order.

Since the system under consideration is a zero position error system, $D(s)$ has a zero at the origin.[5] If $M_1(s)$ also has a pole at the origin, $C(s)$ would have a second-order pole there. The following shows that any pole of $C(s)$ at the origin is of the first order.

If the pole at the origin were of the second order, the inversion integral formula[4] permits $c(t)$ to be written as

$$A_{-2}t+A_{-1}+\sum_{n=0}^{N} B_n \cos(\omega_n t+\phi_n) \tag{5}$$

plus terms of the type $\epsilon^{-a t_k}\cos(\omega t+\phi)$. The last type of term in $c(t)$ corresponds to poles of $C(s)$ in the left half of the complex plane. The second-order pole of $C(s)$ at the origin gives the linear term in $c(t)$ and the other single-order poles of $C(s)$ on the imaginary axis give the cosine terms of $c(t)$.

After a sufficiently long time, the linear term of $c(t)$ will be so large in comparison with the other terms that the amplifier will saturate. Then, the output of the amplifier will be equal in magnitude to e_s and will have the same sign as the error $e(t)$. Since $e(t)=-c(t)$, we have, for t sufficiently large

$$m_1(t)=e_s\frac{e(t)}{|e(t)|}=-e_s\frac{c(t)}{|c(t)|}$$

The linear term in $c(t)$ will give the term A_{-2}/s^2 in the Laurent expansion of $C(s)$ about the origin. The transform of $m_1(t)$ for t sufficiently large will be

$$M_1(s)=-e_s\frac{c(t)}{|c(t)|}\frac{1}{s}$$

One term in the partial fraction expansion of the right member of equation 4 will contain the term

$$\frac{1}{s^2}\lim_{s\to 0}\frac{s\{I(s)+KBN(s)M_1(s)\}}{\frac{D(s)}{s}}$$

$$=-KBe_s\frac{c(t)}{|c(t)|}\frac{1}{s^2}$$

since the constant terms of $N(s)$ and $D(s)/s$ are unity. Then one term of $C(s)$ will be, if we write

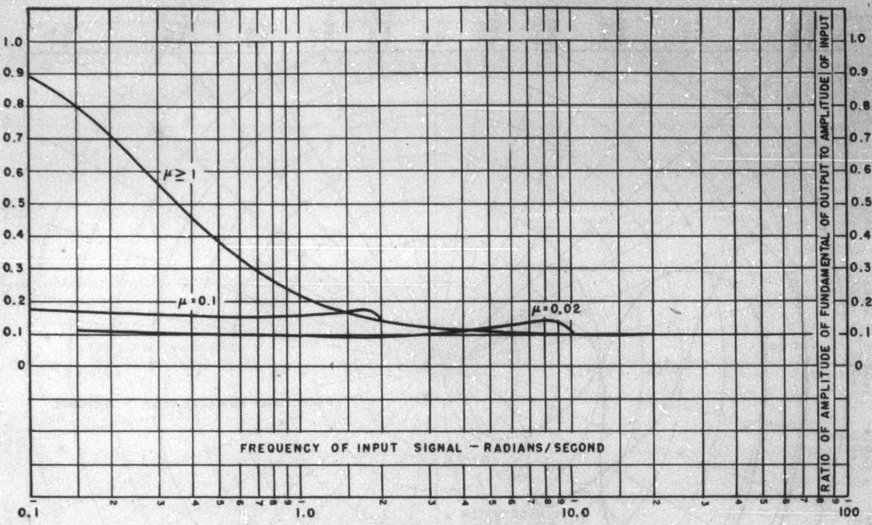


Figure 7. Amplitude characteristics of nonlinear filter for sinusoidal input

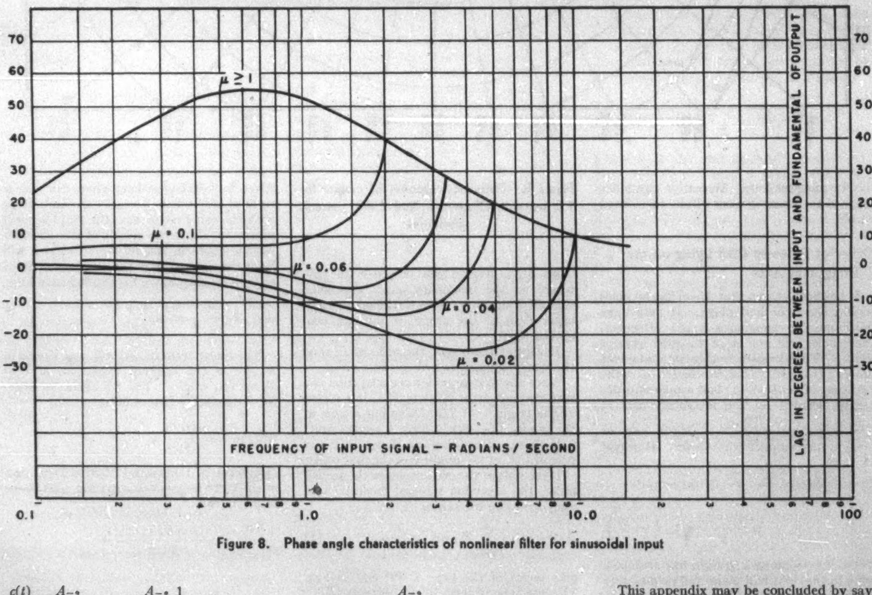


Figure 8. Phase angle characteristics of nonlinear filter for sinusoidal input

$$\frac{c(t)}{|c(t)|}=\frac{A_{-2}}{|A_{-2}|}, \ -KB\,e_s\frac{A_{-2}}{|A_{-2}|}\frac{1}{s^2}$$

From equation 5, one term of $C(s)$ is A_{-2}/s^2. In order that $s^2C(s)$ may be regular at the point $s=0$, we must have

$$A_{-2}=-KB\,e_s\frac{A_{-2}}{|A_{-2}|}$$

where K, B, and e_s are positive. Then $A_{-2}=0$, or there is no second-order pole at the origin.

This appendix may be concluded by saying that the poles of $C(s)$ lie within the left half plane or upon the imaginary axis if the system contains no unstable inner loops. For a zero error system, the poles on the imaginary axis are of the first order.

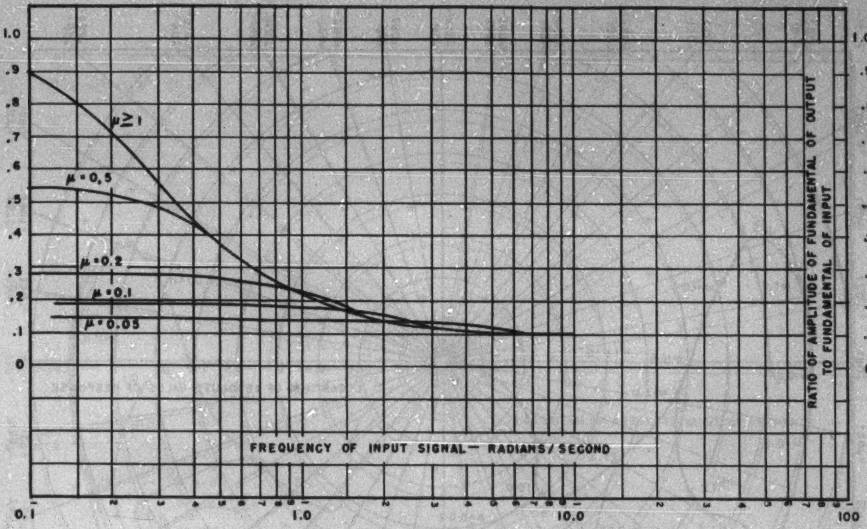


Figure 9. Amplitude characteristics of nonlinear filter for square wave input

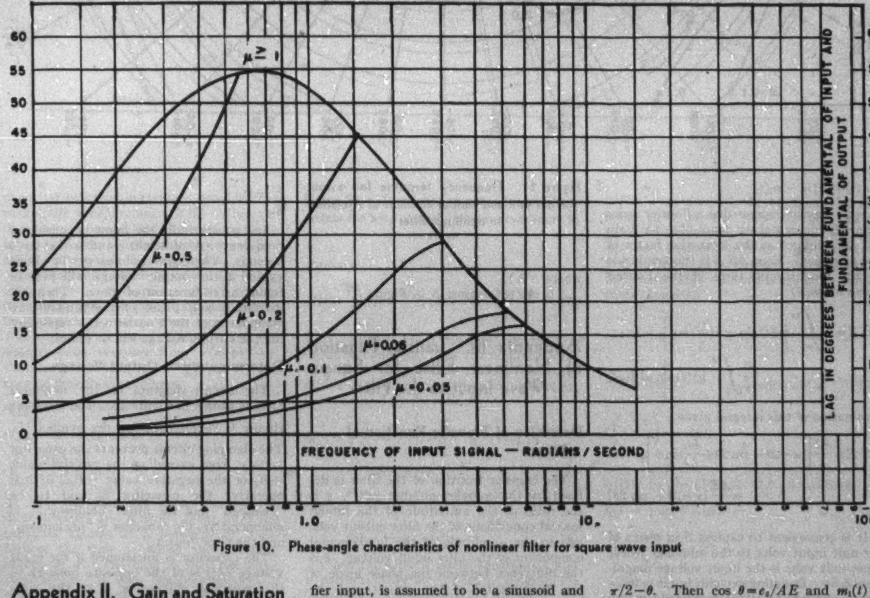


Figure 10. Phase-angle characteristics of nonlinear filter for square wave input

Appendix II. Gain and Saturation Factor of Voltage Amplifier

The saturation characteristic of the amplifier is shown in Figure 2. $e(t)$, the amplifier input, is assumed to be a sinusoid and is written $E \sin \omega t$. $m_1(t)$, the amplifier output, is equal to $AE \sin \omega t$ for the input magnitude less than e_s/A. For larger input magnitudes, $m_1(t) = \pm e_s$. Let the value of ωt at which $E \sin \omega t = e_s/A$ be denoted by $\pi/2 - \theta$. Then $\cos \theta = e_s/AE$ and $m_1(t)$ is given by

$$m_1(t) = AE \sin \omega t \quad \text{for} \quad 0 \leq \omega t \leq \pi/2 - \theta \quad \text{and} \quad \pi/2 + \theta < \omega t \leq \pi$$

$$m_1(t) = e_s \quad \text{for} \quad \pi/2 - \theta < \omega t \leq \pi/2 + \theta$$

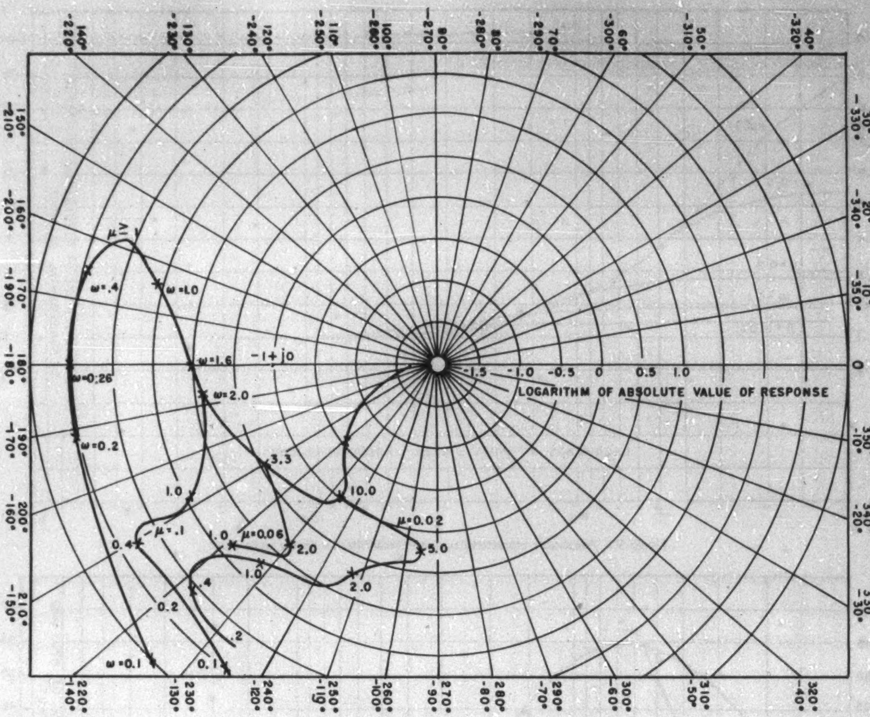


Figure 11. Frequency response for output ≤ 1 per unit and various amounts of saturation in nonlinear filter

$$m_1(t+\pi/\omega) = -m_1(t)$$

$m_1(t)$ may be expressed as a Fourier series whose fundamental we denote by $ASE \sin \omega t$. S is defined as the saturation factor of the amplifier. Since ASE is the amplitude of the fundamental term of the Fourier series for $m_1(t)$

$$ASE = \frac{1}{\pi}\int_0^{2\pi} m_1(t) \sin \omega t \, d(\omega t) = \frac{2}{\pi}\int_0^{\pi} m_1(t) \sin \omega t \, d(\omega t)$$

Evaluation of this integral gives

$$ASE = \frac{AE}{\pi}\left[\pi - 2\theta - \sin 2\theta + \frac{4e_s}{Ae}\sin\theta\right] = \frac{AE}{\pi}[\pi - 2\theta + \sin 2\theta]$$

It is convenient to express S in terms of per-unit input volts to the amplifier where 1 per-unit volts is the input voltage magnitude e_s/A. Denoting per-unit input voltage by N, the input voltage amplitude in per unit is $EA/e_s = N = 1/\cos\theta$. Then

$$S = \left[1 - \frac{2}{\pi}\cos^{-1}\left(\frac{1}{N}\right) + \frac{2}{\pi}\frac{\sqrt{N^2-1}}{N^2}\right]$$

where $N \geq 1$.

S is plotted against N in Figure 5.

Appendix III. Transfer Function of Nonlinear Filter for Square Wave Input to the Filter

Definition of Transfer Function of Filter

The transfer function of the filter is defined by the complex number $g\epsilon^{-i\theta}$. g is the ratio of the amplitude of the fundamental component of the filter output voltage to the amplitude of the fundamental component of the filter input voltage. θ is the difference between the phase angle of the fundamental of the output and the phase angle of the fundamental of the input. If the fundamental of the input is written as $g_1 \sin \omega t$ and the fundamental of the output as $g_2 \sin(\omega t - \theta)$, the transfer function is

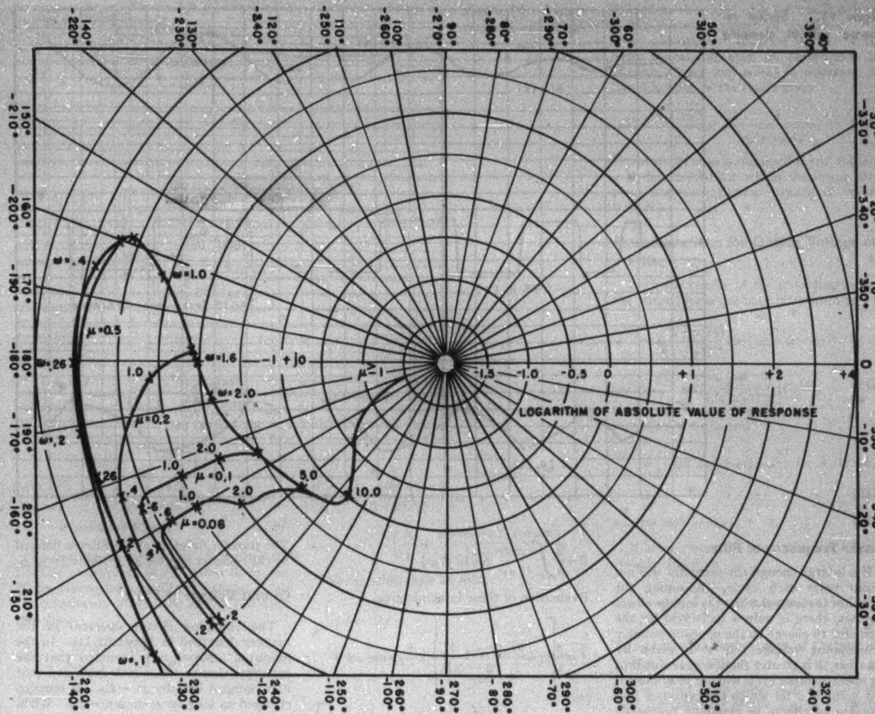
$$\frac{g_2}{g_1}\epsilon^{-i\theta}$$

In this appendix the input is considered a square wave of height e and period $2\pi/\omega$ seconds. Then, $g_1 = 4e \sin \omega t/\pi$. To obtain g_2 and θ the output voltage will first be found as a function of time. Then the amplitude and phase of the fundamental component of the Fourier series representing the output voltage will be found.

Determination of Output Voltage

The circuit diagram of the nonlinear filter is shown in Figure 15. The clamping circuit is represented by the symbol $\boxed{L}$. The clamping circuit prevents the capacitor voltage from exceeding the positive value $+E_s$ or the negative value $-E_s$. If it is operative the capacitor is said to be "clamped" and the filter "nonlinear." If nonoperative, the capacitor is "unclamped" and the filter "linear."

The capacitor is unclamped if the input voltage $e_i(t)$ is of the opposite polarity to the capacitor voltage or is of the same polarity as the capacitor voltage but less than E_s in magnitude. If the capacitor voltage is E_s in magnitude and $e_i(t)$ is of the same polarity as the capacitor voltage and greater in magnitude than E_s, the



capacitor will remain clamped provided the magnitude of $e_i(t)$ remains $\geq E_s$. When the capacitor is unclamped its voltage as a function of time is determined by the differential equations of the linear filter.

In analyzing the case when the filter clamps at a positive and a negative voltage once each cycle, the capacitor is assumed to be clamped at $-E_s$. Then $e_i(t) = -e$. The capacitor remains clamped until the square wave reverses from $-e$ to $+e$. Call this time $t=0$. Then the input voltage may be described as follows

$$e_i(t) = +e \quad 0<\omega t<\pi$$

$$e_i(t) = -e \quad \pi<\omega t<2\pi$$

$$e_i\left(t+\frac{2\pi}{\omega}\right) = -e_i(t)$$

At $t=0$ the capacitor unclamps and the filter becomes and remains linear until the capacitor clamps at a potential of $+E_s$. Denote the time at which the clamping occurs by Ψ. The filter is then nonlinear until $\omega t=\pi$. It then becomes linear and remains linear until $t=\pi/\omega+\Psi$ where the capacitor clamps at the potential $-E_s$ and remains clamped until $\omega t=2\pi$. Then this cycle is repeated.

Referring to Figure 15 and employing the superposition theorem equation 6 is

Figure 12. Frequency response for nonlinear system for S=1.0 and square wave input to filter

obtained

$$e_0(t) = \frac{T_2}{T_1}e_i(t) + \frac{\delta T}{T_1}e_c(t) \tag{6}$$

This equation holds whether the filter is linear or nonlinear. For the nonlinear case $e_i(t)$ has the magnitude e and $e_c(t)$ the magnitude E_s. Then, for the nonlinear filter, if μ is defined by the ratio E_s/e

$$e_0(t) = +\frac{T_2}{T_1}e + \frac{\delta T}{T_1}E_s = e\left[1-\frac{\delta T}{T_1}(1-\mu)\right]$$
$$\text{for } \omega\Psi<\omega t<\pi \tag{7}$$

$$e_0(t) = -\frac{T_2}{T_1}e - \frac{\delta T}{T_1}E_s = -e\left[1-\frac{\delta T}{T_1}(1-\mu)\right]$$
$$\text{for } \pi+\omega\Psi<\omega t<2\pi \tag{8}$$

When the filter is linear equations 9 and 10 are obtained from the circuit diagram for $0<\omega t<\omega\Psi$

$$e_i(t) = (R_1+R_2)i(t) + e_c(t) \tag{9}$$

$$e_c(t) = -E_s + \frac{1}{C}\int_0^t i(t)\,dt \tag{10}$$

For $\pi<\omega t\leq\pi+\omega\Psi$ the equations for the linear filter are the same except $-E_s$ of equation 10 is replaced by $+E_s$. To obtain $e_0(t)$ as a function of time equations 6, 9, and 10 are solved by Laplace transformation methods. This solution is given by equations 11 and 12

$$e_0(t) = +e\left[1-\frac{\delta T}{T_1}(1+\mu)e^{-\frac{t}{T_1}}\right] \quad 0<\omega t\leq\omega\Psi$$

$$e_0(t) = +e\left[1-\frac{\delta T}{T_1}(1-\mu)\right]\omega\Psi\leq\omega t<\pi \tag{11}$$

$$e_0\left(t+\frac{\pi}{\omega}\right) = -e_0(t)$$

$$e_c(t) = +e\left(1-(1+\mu)\epsilon^{-\frac{t}{T_1}}\right) \quad 0\leq\omega t\leq\omega\Psi$$

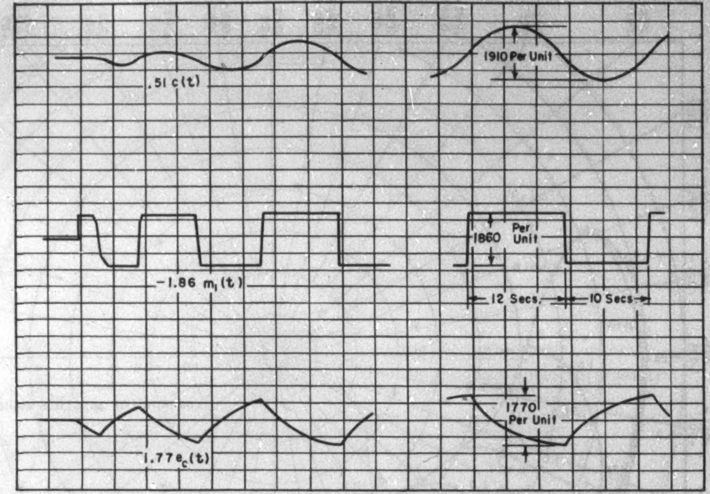
$$e_c(t) = +\mu e\omega\Psi\leq\omega t\leq\pi \tag{12}$$

$$e_c\left(t+\frac{\pi}{\omega}\right) = -e_c(t)$$

The value of Ψ is determined as follows. For $t=\Psi$, $e_c(t)=\mu e$. Putting $e_c(t)=\mu e$ in the first expression of equation 12 we obtain equation 13

$$\epsilon^{-\frac{\Psi}{T_1}} = \frac{1-\mu}{1+\mu} \tag{13}$$

Figure 13. System response without clamping



Linear Frequency of Filter

If ω is large enough the capacitor will not clamp twice each cycle. Assuming the capacitor is clamped when the input voltage reverses, there is only a half-cycle for the capacitor to charge to the opposite polarity of clamping voltage. If Ψ as given by equation 13 is greater than π/ω, two clamping operations per cycle will not be obtained. The value of ω for which the capacitor just manages to charge to clamping potentials twice a cycle is denoted by ω_l. It is given by $\omega_l\Psi=\pi$ or

$$\epsilon^{-\frac{\pi}{\omega_l T_1}}=\frac{1-\mu}{1+\mu} \tag{14}$$

For $\omega \geq \omega_l$, the transfer function of the filter is the same as the transfer function of the filter without its clamping circuit. Only for $\omega<\omega_l$ does the transfer function of the filter depend upon the clamping circuit.

Fundamental Component of Filter Output

The fundamental component of output voltage is written as $eA \sin \omega t+eB \cos \omega t=g_2 \sin(\omega t+\theta)$. Then

$$\frac{g_2}{e}=\sqrt{A^2+B^2} \tag{15}$$

$$\tan^{-1}\theta=\frac{B}{A} \tag{16}$$

A is given by the equation

$$A=\frac{1}{\pi}\int_0^{2\pi}\frac{e_0(t)}{e}\sin \omega t\, d(\omega t)=$$

$$\frac{2}{\pi}\int_0^{\pi}\frac{e_0(t)}{e}\sin \omega t\, d(\omega t)$$

and B is given by

$$B=\frac{2}{\pi}\int_0^{\pi}\frac{e_0(t)}{e}\cos \omega t\, d(\omega t)$$

Evaluation of these integrals gives

$$A=\frac{4}{\pi}\left[1-\frac{\delta T}{T_1}\frac{1-\mu}{2}-\frac{\delta T}{2T_1}\left\{(1-\mu)\cos \omega\Psi+\frac{1+\mu-(1-\mu)\left(\cos \omega\Psi+\frac{1}{\omega T_1}\sin \omega\Psi\right)}{1+\left(\frac{1}{\omega T_1}\right)^2}\right\}\right] \tag{17}$$

$$B=\frac{2}{\pi}\frac{\delta T}{T_1}\left[(1-\mu)\sin \omega\Psi-\frac{\frac{1+\mu}{\omega T_1}-(1-\mu)\left(\frac{\cos \omega\Psi}{\omega T_1}-\sin \omega\Psi\right)}{1+\left(\frac{1}{\omega T_1}\right)^2}\right] \tag{18}$$

The transfer function is plotted in Figures 9 and 10 for various values of μ.

Appendix IV. Transfer Function of Nonlinear Filter for Sine Wave Input to the Filter

Definition of Transfer Function of Filter

The input to the filter is written as $g_1 \sin \omega t$, the output of the filter as $g_2 \sin(\omega t-\theta)$. The transfer function of the filter is defined by the complex number $g\epsilon^{-i\theta}$ where $g=g_1/g_2$.

Output Voltage of Filter

The clamping circuit operates in the manner described in Appendix III. In the following analysis it is assumed that the capacitor clamps twice each cycle and that it is clamped initially at $-E_S$. It remains clamped as long as $g_1 \sin \omega t<-E_S$. When $g_1 \sin \omega t=-E_S$ and is increasing, the filter becomes linear. Calling this value of time β, we have

$$\sin \omega\beta=-\frac{g_1}{E_S}=-\mu \text{ and } \cos \omega\beta=\sqrt{1-\mu^2}$$

To facilitate the analysis the substitution $t=\beta+v$ is made. The filter is linear for $v\geq 0$ and remains so until such time as the capacitor again clamps. Denote this value of v by Ψ. The filter then is linear for $0\leq v\leq\Psi$. Equations 6, 9, and 10 describe the linear filter when in these equations v is written in place of t.

After clamping at $+E_S$, the capacitor will unclamp when $g_1 \sin \omega t=E_S$ and is decreasing. This occurs at $t=\beta+\pi/\omega$. The filter then will remain linear until $t=\beta+\pi/\omega+\Psi$, at which time it will clamp at $-E_S$. For $\pi/\omega\leq v\leq\pi/\omega+\Psi$, the equations for the linear filter are as given by equations 6, 9, and 10, except that $-E_S$ of equation 10 is replaced by $+E_S$.

The filter will be nonlinear for $\Psi<v<\pi/\omega$ and $\pi/\omega+\Psi<v<2\pi/\omega$. As discussed in Appendix III, equation 6 is valid for this condition. During this time the capacitor voltage is $+E_S$ or $-E_S$ and the input voltage is $g_1 \sin\{\omega\beta+\omega v\}$.

Substituting these values into equation 6 we have

$$\frac{e_0(v)}{g_1}=\frac{T_2}{T_1}\sin\{\omega\beta+\omega v\}+\frac{\delta T}{T_1}\mu\Psi\leq v\leq\pi/\omega \tag{19}$$

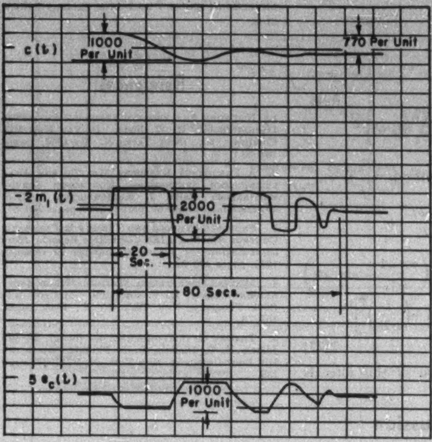


Figure 14. System response with clamping

$$\frac{e_0(v)}{g_1}=\frac{T_2}{T_1}\sin\{\omega\beta+\omega v\}-\frac{\delta T}{T_1}\mu\times$$
$$\pi/\omega+\Psi\leq v\leq 2\pi/\omega \tag{20}$$

To find the output of the linear filter the Laplace transforms of equations 6, 9, and 10 are taken and solved for $E_0(s)$ and $E_c(s)$, the transforms of $e_0(v)$ and $e_c(v)$. The solution gives

$$E_c(s)=\frac{E_i(s)}{1+T_1s}-\frac{T_1E_S}{1+T_1s} \tag{11A}$$

and

$$E_0(s)=\frac{1+T_2s}{1+T_1s}E_i(s)-\frac{\delta T}{T_1}\frac{E_S}{s+\frac{1}{T_1}} \tag{12A}$$

$E_i(s)$ in these equations is the transform of $g_1\sin(\omega\beta+\omega v)$. It is given by

$$E_i(s)=g_1\frac{s\sin\omega\beta+\omega\cos\omega\beta}{s^2+\omega^2}$$

The inverse transforms of $E_0(s)$ and $E_c(s)$ are found to be

$$\frac{e_0(v)}{g_1}=\sqrt{\frac{1+\omega^2T_2^2}{1+\omega^2T_1^2}}\sin(\omega\beta+\omega v-\phi)-$$
$$\left[\frac{\delta T}{T_1}\frac{E_S}{g_1}+\frac{\omega\delta T}{1+\omega^2T_1^2}\left(\frac{\sin\omega\beta}{\omega T_1}-\cos\omega\beta\right)\right]\epsilon^{-\frac{v}{T_1}} \tag{21}$$

where $\tan\phi=\omega\delta T/1+\omega^2T_1T_2$.

$$\frac{E_c(v)}{g_1}=\frac{\sin(\omega\beta+\omega v-\delta)}{\sqrt{1+\omega^2T_1^2}}-$$
$$\left(\frac{E_S}{g_1}+\frac{\sin\omega\beta-\omega T_1\cos\omega\beta}{1+\omega^2T_1^2}\right)\epsilon^{-\frac{v}{T_1}} \tag{22}$$

where $\tan\delta=\omega T_1$.

The filter is linear until $v=\Psi$ at which time the capacitor clamps at $+E_S$. Ψ is the solution of

$$\frac{E_S}{g_1}=\frac{\sin(\omega\beta+\omega\Psi-\delta)}{1+\omega^2T_1^2}-$$
$$\left(\frac{E_S}{g_1}+\frac{\sin\omega\beta-\omega T_1\cos\omega\beta}{1+\omega^2T_1^2}\right)\epsilon^{-\frac{\Psi}{T_1}}$$

This equation may be written

$$\mu\left(1+\epsilon^{-\frac{\Psi}{T_1}}\right)=$$
$$\frac{\sin(\omega\beta+\omega\Psi-\delta)-\sin(\omega\beta-\delta)\epsilon^{-\frac{\Psi}{T_1}}}{\sqrt{1+\omega^2T_1^2}} \tag{23}$$

Seemingly, equation 23 is best solved for Ψ by a cut-and-try method.

Linear Frequency of Filter

In the preceding analysis it has been assumed that the capacitor is clamped during a part of the negative half cycle of the input voltage $g_1\sin(\omega\beta+\omega v)$. If it is to clamp during the succeeding positive half cycle, it must do so for $\omega\Psi\leq\pi$ radians. ω_l is defined as the value of ω for which $\omega\Psi=\pi$, as the linear frequency of the filter. For $\omega>\omega_l$, the transfer function is that of the linear filter. ω_l may be determined by putting $\omega\Psi=\pi$ and $\omega=\omega_l$ in equation 23. Simplification of the result gives

$$\omega_lT_1=\frac{\sqrt{1-\mu^2}}{\mu}\text{ or }\mu=\frac{1}{\sqrt{1+\omega_l^2T_1^2}} \tag{24}$$

Equation 24 also is obtained if the value of ω is determined for which the linear filter gives a peak capacitor voltage of magnitude E_S.

Final Equations for Output Voltage of Filter

With the results of the preceding section the output for $\omega<\omega_l$ may be written as

$$\frac{e_0(v)}{g_1}=\sqrt{\frac{1+\omega^2T_2^2}{1+\omega^2T_1^2}}\sin(\omega\beta+\omega v-\phi)-$$
$$\frac{\delta T}{T_1}\left[\mu+\frac{\sin(\omega\beta-\delta)}{\sqrt{1+\omega^2T_1^2}}\right]\epsilon^{-\frac{v}{T_1}} \tag{25A}$$

for $0\leq v\leq\Psi<\pi/\omega$ and

$$\frac{e_0(v)}{g_1}=\frac{T_2}{T_1}\sin\{\omega\beta+\omega v\}+\mu\frac{\delta T}{T_1}\text{ for }\Psi\leq v\leq\pi/\omega \tag{25B}$$

$$e_0(v+\pi/\omega)=-e_0(v) \tag{25C}$$

If $\omega>\omega_l$, we have from the linear filter transfer function

$$\frac{e_0(v)}{g_1}=\sqrt{\frac{1+\omega^2T_2^2}{1+\omega^2T_1^2}}\sin(\omega\beta+\omega v-\phi)$$

Fundamental Component of Output of Nonlinear Filter

The input to the filter is $g_1\sin(\omega\beta+\omega v)$. The fundamental of the output is written as $g_2\sin(\omega v+\theta_2)=g_1A\sin\omega v+g_1B\cos\omega v$. Then $g_2/g_1=\sqrt{A^2+B^2}$ and $\tan^{-1}\theta_2=B/A$.

A is given by the equation

$$A=\frac{1}{\pi}\int_0^{2\pi}\frac{e_0(v)}{g_1}\sin\omega v\,d(\omega v)=$$
$$\frac{2}{\pi}\int_0^{\pi}\frac{e_0(v)}{g_1}\sin\omega v\;d(\omega v)$$

Similarly B is given by

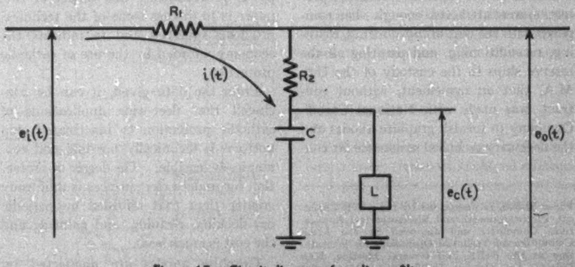


Figure 15. Circuit diagram of nonlinear filter

$T_1=(R_1+R_2)C=5$
$T_2=R_2C=0.5$
$\delta T=T_1-T_2=4.5$

$$B=\frac{2}{\pi}\int_0^\pi \frac{e_0(v)}{g_1}\cos \omega v\, d(\omega v)$$

Using equation 25 the evaluation of these integrals gives

$$\pi A=\pi\frac{T_2}{T_1}\cos\omega\beta+\omega\Psi \left[r_l\cos(\omega\beta-\phi)-\frac{T_2}{T_1}\cos\omega\beta\right]- \sin\omega\Psi\left[r_l\cos(\omega\Psi+\omega\beta-\phi)-\frac{T_2}{T_1}\times \cos(\omega\Psi+\omega\beta)\right]+ \frac{2\mu\delta T(\omega_l-\omega)}{T_1\omega}\frac{1-\epsilon^{-\frac{\Psi}{T_1}}\left(\frac{\sin\omega\Psi}{\omega T_1}+\cos\omega\Psi\right)}{\left[1+\left(\frac{1}{\omega T_1}\right)^2\right]^2}+ \frac{2\mu\delta T}{T_1}(1+\cos\omega\Psi) \qquad (26)$$

where $r_l=\sqrt{\dfrac{1+\omega^2T_2^2}{1+\omega^2T_1^2}}$

$$\pi\beta=\pi\frac{T_2}{T_1}\sin\omega\beta+\omega\Psi \left[r_l\sin(\omega\beta-\phi)-\frac{T_2}{T_1}\sin\omega\beta\right]+ \sin\omega\Psi\left[r_l\sin(\omega\Psi+\omega\beta-\phi)-\frac{T_2}{T_1}\times \sin(\omega\Psi+\omega\beta)-\frac{2\mu\delta T}{T_1}\right]+ \frac{2\mu\delta T(\omega_l-\omega)}{T_1\omega}\frac{\frac{1}{\omega T_1}-\epsilon^{-\frac{\Psi}{T_1}}\left(\frac{\cos\omega\Psi}{\omega T_1}-\sin\omega\Psi\right)}{\left[1+\left(\frac{1}{\omega T_1}\right)^2\right]^2} \qquad (27)$$

The transfer function of this filter is plotted in Figures 7 and 8 for various values of μ.

References

1. Network Analysis and Feedback Amplifier Design (book), **H. W. Bode.** D. Van Nostrand Company, Inc., New York, N. Y., 1945, pages 162–63.

2. A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms, **Ralph J. Kochenburger.** *AIEE Transactions*, volume 69, part I, 1950, pages 270–84.

3. Proposed Symbols and Terms for Feedback Control Systems, **AIEE Committee Report.** *Electrical Engineering*, volume 70, number 10, October 1951, pages 905–09.

4. Modern Operational Mathematics in Engineering (book), **R. V. Churchill.** McGraw-Hill Book Company, New York, N. Y., 1944, pages 167–72.

5. Theory of Servomechanisms (book), **H. M. James, N. B. Nichols, R. S. Phillips.** McGraw-Hill Book Company, New York, N. Y., 1947, page 139.

No Discussion

The Use of Graphite Anodes for Cathodic Protection of the Bottoms of Inactivated Ships

J. P. OLIVER
ASSOCIATE MEMBER AIEE

THE FIRST large scale experimental application of graphite anodes and an impressed current to cathodically protect the bottoms of inactive ships was planned by the United States Maritime Administration (U.S.M.A.) in 1949 as a result of discussions with National Carbon Company representatives.

The economic aspects of cathodic protection were attractive enough when compared with the cost of dry-docking, cleaning, reconditioning, and painting all the reserve ships in the custody of the U.S. M.A. that an agreement, without contract, was made with National Carbon Company to furnish graphite anodes and the necessary technical assistance for conducting tests on ships to be designated by the U.S.M.A.

The first actual installation of an impressed current system using rectifiers and graphite anodes was made in January 1950 and was so successful that the program was extended to include experimental applications at all fleet sites where power is available. The object of this paper is to discuss some of the technical problems involved and to indicate the economy afforded by the use of cathodic protection.

From the data given, it can be concluded that fleet-wide applications of cathodic protection to inactivated ship bottoms is technically possible and economically feasible. The degree of protection for underwater surfaces is uniformly greater than that afforded by periodic dry-docking, cleaning, and painting and the cost is much less.

Graphite anodes are unaffected by either salt or fresh water; neither are they subject to chemical deterioration when not in use. At the same time the current discharged is many times the theoretica electrochemical equivalent of 4,042 ampere hours per pound. These factors contribute to the success of the systems which the Maritime Administration has so carefully developed.

Not included in this report are tests now under way which fully demonstrate that groups of ships as well as individual vessels can be protected by grids of anodes, either remotely located or lying under the keels. Rectifier units may be placed aboard the ships but are preferably located on shore. Such an arrangement provides economical control from points on shore while reducing to a minimum any inconvenience when placing the ship in active service. By placing anodes on the

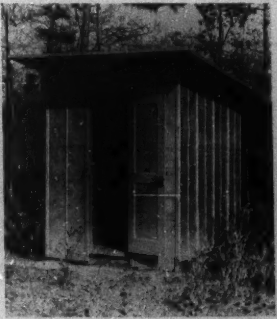
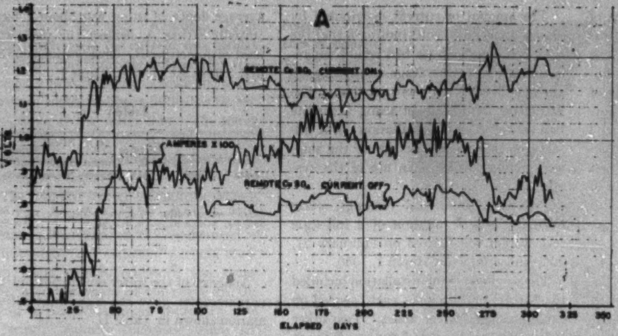


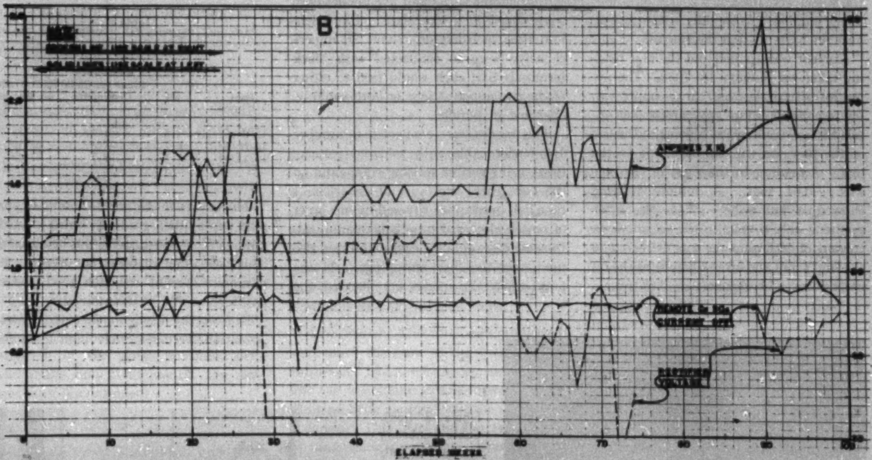
Figure 1. Rectifier shack in swamp area

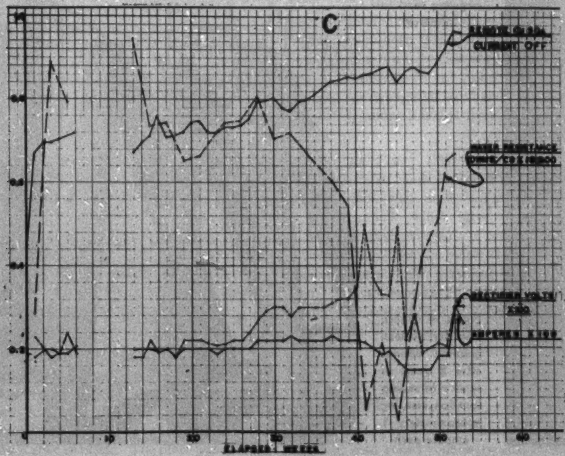
Paper 52-306, recommended by the AIEE Chemical, Electrochemical, and Electrothermal Applications Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted June 11, 1952; made available for printing September 8, 1952.

J. P. Oliver is with the National Carbon Company, Cleveland, Ohio.

Figure 2. Summary of data for test numbers 1, 2, and 3







bottom and rectifiers on shore, the system is readily available for use with any vessel that may be moored nearby. This is particularly desirable in the case of the reserve ships in custody of the Maritime Administration where vessels are reactivated from time to time, to be returned when they are again taken out of active service.

Test Number 1

The first installation was made on a former troop transport anchored adjacent to a pier that was an extension of the causeway extending approximately 1,000 feet from the shore. The vessel was 684 feet long, had a beam of 74 feet, and a draft of 32 feet. The water in which she was anchored had an average depth of 39 feet, with approximately 7 feet of water under the keel at high tide, and approximately 3 feet at low tide. The area of submerged metal to be protected was about 65,000 square feet. Because the vessel had not been dry-docked for approximately 2 years, there was no way of determining the area of underwater surface still covered by paint.

The initial plan was to place the graphite anodes vertically along the shore just above high tide level, but digging holes in any area near the causeway was found to be practically impossible; therefore, six 4 by 40-inch graphite anodes were placed in shallow water on the side of the causeway opposite and about 900 feet from the ship. Because the rectifiers had not yet been delivered, storage batteries aboard the vessel were used as a source of power, using number 4/0 welding cable to connect the batteries and the anodes.

Current was applied and the first reading taken on January 18, 1950. The potential of the ship hull to a copper sulphate half cell before any current was applied was −0.45 volt at a point to the left of the bow, on the side adjacent to the pier. A reading to the right of the bow was −0.64 volt. A remote reading showed the hull to be −0.645 volt. After applying only 6 volts to the six anodes a current of 20.6 amperes was delivered to the ship's hull and after 30 minutes the remote potential had risen to −0.67 volt.

The amount of current distributed by the six anodes at 8 volts indicated that the use of a remote anode bed would not require an excessive voltage and that good current distribution was probable. (The attractive advantage of using a remote bed included lower installation costs, absence of cables over the side of the ship, and greater flexibility in use of the system.)

Following the initial readings, using batteries, the installation of a full-scale test system was started on February 27, 1950 and completed March 1, 1950.

The original anode installation included two strings of 14 anodes each, spaced on cable headers 15 feet apart. Two other strings having seven and five anodes each were added at a later date. Two selenium type rectifiers of 12-volt 80-ampere capacity were so connected with block terminals that each would feed to either of the two anode strings or both could be operated in parallel.

The anodes were connected to number 4/3 feeder cable with soldered connections, using rubber splicing compound on the bare wire, one layer of electrician's tape half overlapped covered by one layer of scotch electrical tape half overlapped. In some cases splicing compound was also used over the tape wrappings. At the point where the anodes were located, the average depth of water was about 7 feet. Initially string *A*, nearest the pier, was used with rectifier *B*. Observations on the first 2 days are given in Table I.

Table I

Date	Time	Rectifier B, Volts	D-C Output, Amperes	Anode String	Potential ($CuSO_4$), Volts
3/1/50	3:30 p.m.	10.7	49.0	A	−0.86
3/2/50	8:30 a.m.	10.5	46.0	A	−0.86
	3:30 p.m.	10.5	46.0	A	−0.88

Analyses of the water on the bottom on various dates gave the significant information shown in Table II.

Table II

Date	pH	Electrica Resistance, Ohms per Cubic Centimeter	NaCl Grains per Gallon
12/8/49	6.5	61.0	67.81
12/11/49	6.7	61.3	67.15
12/15/49	7.1	58.7	64.42

The charts in Figure 2 show that current was applied gradually until a maximum of 90 to 100 amperes was being used. This was done in order to observe the minimum current at which good protective levels could be maintained.

Noteworthy also is the fact that remote potentials were relatively high from the

Figure 3 (left). Cable connecting lug on ship hull, negative lead to rectifier

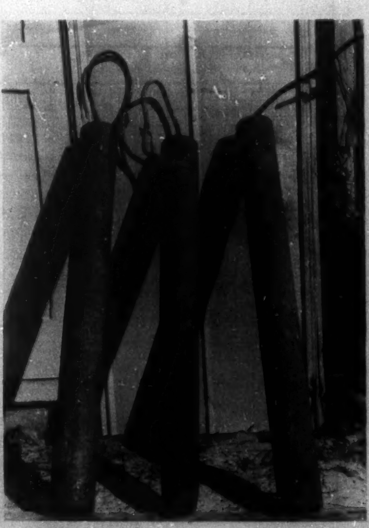


Figure 4 (right). Four by 40-inch graphite anodes after 2 years of service in swamp location, carrying approximately 1 ampere each beginning of the test. Readings taken immediately after the current was turned off proved that complete polarization was necessary because of heavy scale and paint coverage which had been indicated by observations at the water line. The lower line on Figure 2(A) indicates polarization while the upper line indicates polarization plus overvoltage.

The trial was started officially on March 1. The conductivity of the water increased as the summer season advanced as indicated by the increased current readings while voltage remained the same. The small variations on the curve are the result of salinity changes caused by rain or wind. Salinity increased when sea water was blown into the bay; rains decreased the salinity because of the high flow of fresh water into the bay.

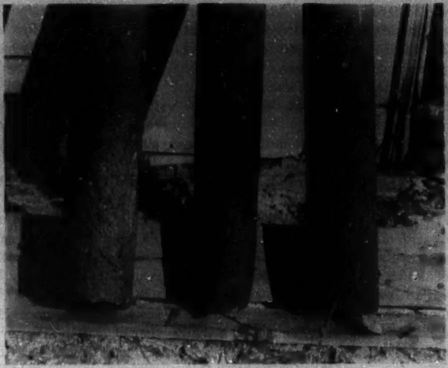


Figure 5. Lower ends of National 4 by 40-inch graphite anodes

TEST COUPONS

Because of inability to observe directly the results on the ship hull, an effort was made to evaluate coupons hung over the side and under the keel, both in and out of circuit. The coupons were foot square cleaned steel plates drilled to accomodate suspension and electric connections. The plates were installed in January 1950, but current was not applied until March 1. Potential readings and examination of the coupons later in March 1950 indicated that only trends could be established. Inconsistencies between potentials and condition of the plates precluded definite results and consequently the coupons were discontinued.

WELDED PLATES

On April 5, two steel plates were thoroughly sand-blasted and welded 5 feet apart at right angles to the hull at the water line. The portions of the plates above the water line were brush-coated with red oxide and oil. On April 20, another plate, 12 by 30 inches, was similarly installed but electrically insulated from the hull with 9 inches submerged.

Examination of the plates on May 15, disclosed a thin film below the water line that was easily removed, leaving a clean surface. Areas subjected to maximum wave action were roughened, and the accumulated film was more difficult to remove. Another examination was made on October 19, 1950, after the plates had been under test more than 6 months.

Forward Plate: The surface of the forward side under the water was covered with a heavy growth of brown algae which was removed with a stiff bristle scrub brush. The exposed surface had a general appearance of light gray areas mixed with dark gray to black. The plate was smooth to the touch, with a few barnacles but no pits or rust nodules. A very few small reddish spots could be seen just below the water line. There was a heavy rust formation to a height of about 7 inches above the water line, appearing as a scale about 1/32-inch thick in some places, while above this the plate was still covered with the red oxide mixture and was in good condition. When observed again on June 14, 1951, the rust band extended upward only an additional 1/2 inch.

After Plate: The after side appeared very much the same as the forward side. Both sides had the same appearance below the water line as the forward plate, showing little change from the earlier inspection. This plate did not receive the oil and oxide treatment. The band of rust, approximately 1/32-inch thick in spots, extended about 9 inches above the water line on both sides. Above the rust band there was a light coat of rust appearing as a discoloration. When first inspected the heavy rust band extended upward only about 2 inches from the water line.

Insulated Plate: The insulated panel was very rough below the water line with a heavy black rust scale, approximately 1/16-inch thick, which lifted off easily. The black rust clouded the water when brushed off; after the water cleared, the plate appeared light gray but badly pitted. Above the water line the plate was rough and rusty about 9 inches upward; above this there was a light coat of rust similar to that on the unpainted after plate.

SUSPENDED RODS

On May 9, two 1-inch diameter by 12-$^1/_2$ feet cold rolled rods were suspended vertically from the hull so that 11 feet of each was immersed. One rod was out of circuit while the other was electrically connected to the hull by number 00 welding cable. Rods were used to reduce edge effects to a minimum.

The rod was in circuit, and thus protected, was in excellent condition when inspected on May 15, showing no rust over its entire submerged length.

Another examination was made on October 19, 1950, after more than 5 months of testing.

The protected rod was covered with a growth of slime under which the rod was in the same condition as when previously inspected. There were a number of brown spots not easily removed, probably caused by raising the rod and cleaning for observation by visitors. One or two places that had been painted with glyptol remained bright. Rust extended about 7 inches above the water line.

The underwater portion of the unprotected rod was very rough and covered with rust nodules and pits. A red rust on the surface with a black oxide underneath covered the entire rod, exposing a bright gray but deeply pitted surface when brushed off. The rust extended about 9 inches above the water line. The rust above the water line was harder, more firm, and more difficult to remove.

SUSPENDED PLATE

Since the results of the test coupons were so inconclusive, a large steel plate, 43 by 39 inches was suspended on May 11. A 2-inch strip on each side was painted with glyptol to provide areas for comparison. Number 00 cable was used to permit raising and lowering for complete examination.

The first examination, on May 15, revealed that about one quarter of the surface on both sides was covered by a light rust which was easily removed, the plate appearing dark in this area. The balance of the plate was bright and clean, while the line of protection at the water line was well defined.

The plate was examined again on October 24, after more than 5 months of testing. On the forward side below the water line a reddish tinge covered most of the surface, some light and some darker red, with original light gray color remaining in spots. There were no pits or rust nodules visible, but the plate was covered with a heavy marine growth and a large number of barnacles. The growth was much more difficult to remove than that on the welded plates.

The after side of the plate had a similar appearance below the water line except for a greater percentage of gray surface; it was smooth to the touch with no pits or active rust spots.

The plate was covered with rust above the water line, the scale being 1/32-inch thick in some spots. The forward side had been coated with red oxide which later had been removed and another oxide mixture applied. Very little of the coating remained except at the top of the plate.

CONCLUSIONS

The test was abruptly terminated on January 16, 1951, when the pier was destroyed by fire originating on the causeway necessitating removal of the ship to another site. The evidence to sustain the indications that the ship had been adequately protected at all submerged points included readings taken at 4-foot intervals by keel-hauling a half cell at midpoints forward and aft, as well as remote half-cell readings. Plate and rod observations have been described. This test was not complete in itself, but provided guidance for a number of others. All equipment was removed to another location.

The four anode strings were recovered, inspected, and calipered individually. Based on a nominal diameter of 4 inches, the conditions of the anodes was found to be:

String *A*—Ten anodes, 1/8-inch average apparent loss; four anodes, no appreciable loss.

String *B*—Eleven anodes, 1/16-inch average apparent loss; three anodes, no appreciable loss.

String *C*—Seven anodes, 1/8-inch average apparent loss; 1 anode, no appreciable loss.

String *D*—Five anodes, 1/4-inch average apparent loss, of these, one lost 1/2 inch, another 5/16 inch.

All anodes appeared to lose little if any at the center, the loss being at the ends, greatest at the ends having the greatest volume of water through which to discharge current; anode strings losing the most were those discharging the most current. String *D* and a part of string *A* were under the pier on a rocky bottom. Strings *B* and *C* were sunk in 1 or 2 feet of mud, discharging less current and losing less weight.

The anodes were in use $10^{1}/_{2}$ months. Each had an original surface area of $3^{1}/_{2}$ square feet and carried an average current of approximately $2^{1}/_{4}$ amperes, or 0.64 ampere per square foot. The losses sustained were sufficiently low to indicate that at least a 10-year life could be expected in this service. The data are summarized in Figure 2(A).

Test Number 2

The favorable developments of the foregoing trials led to a similar test on a ship located at a fleet site where the water is essentially fresh, being only slightly brackish at infrequent intervals. The ship selected was 425 feet long with a beam of 57 feet. It is farthest upstream in a river, being separated from an adjacent vessel by a 20-foot fender. At low tide the bow is "mudded in," water flowing under about 60 per cent of the keel. Manilla rope was substituted for all wire mooring cables to reduce electric contacts with the adjacent ship to a minimum. The vessel's bow is at the edge of a swamp area, about 500 feet from solid ground where the power source is located. The hull was well rusted below the water line, the scale being approximately 1/2-inch thick at points near the water line.

Original exploration was made July 7 and 8, 1949. In March 1950, plans were made to put in a full-scale test similar to the one already described. Seven 4 by 40-inch graphite anodes were placed in holes on the shore, connected in parallel with two 4 by 40-inch anodes submerged in the swamp about 15 feet from the shore. Using storage batteries the current readings were as follows:

Volts	Amperes
11.5	0.5
17.5	1.0
25.5	1.5
29.5	2.0

Following the initial result with batteries, it was decided to make a full-scale experimental trial.

The public utility current supply was tapped from the fleet service main line at the watchman's shanty and carried to the rectifier about 600 feet away by number 4/3 rubber covered cable. Both 230- and 115-volt alternating current is available at the rectifier station.

Table III

Date	Average Remote $CuSO_4$ Current Off	Rectifier		Water Resistance, Ohm per Cubic Centimeter
		Volts	Amperes	
11/16/50	672	18.5	22.5	2,950
11/17/50	695	18	26	
11/22/50	685	20	20	
11/24/50	695	20	20	
11/28/50	700	20	20	
11/30/50	695	18	20	8,890
12/5/50	715	18	25	
12/8/50	710	19	19	
12/12/50	710	19	21	
12/15/50	710	19	24	7,960
12/21/50	720	21	19	
Test interrupted for 43 days				
2/15/51	675	20	18	9,433
2/21/51	695	20	18	
3/1/51	735	20	22	7,256
3/8/51	760	19	19	
3/15/51	710	20	20	7,429
3/22/51	715	18	18	
3/29/51	720	20	22	6,574
4/5/51	745	20	22	
4/12/51	755	20	22	6,671
4/20/51	720	20	21	
4/26/51	720	19	21	
5/3/51	730	20	21	7,450
5/10/51	730	20	22	
5/17/51	735	20	22	7,468
5/24/51	735	20	24	
5/31/51	790	22	27	8,032
6/7/51	795	22	29	
6/14/51	805	22	30	7,257
6/21/51	780	22	30	
6/28/51	776	23	28	7,401
7/5/51	795	22	30	
7/12/51	800	22	30	6,720
7/19/51	815	22	30	
7/26/51	830	22	30	
8/2/51	842	23	31	6,045
8/9/51	847	22	32	
8/16/51	857	22	32	5,473
8/23/51	855	22	35	
8/30/51	860	21.5	50	552
9/6/51	862	20	37	
9/14/51	875	18.5	33.5	2,187
9/21/51	879	19	33	
Current was off for 12 hours on September 13, 1951				
9/28/51	825	17	49	271
10/4/51	852	15	22	
10/11/51	860	15	28	
10/18/51	850	15	19	4,212
10/25/51	847	15	20	
11/2/51	875	18	21	5,197
12/3/51	905	18	20	6,532
1/2/52	968	30	30	6,761
2/4/52	967	30	26	

The rectifier was a selenium stack type for cathodic protection applications, outdoor oil-immersed model, having a rated capacity for d-c output of 18 amperes at 60 volts, with adjustable voltage features. It was installed on shore, approximately 500 feet straight ahead of the vessel bow. The leads from the rectifier to each anode were stranded number 8 single conductor with polyethylene-type synthetic insulation that is resistant to water and oil, and may be either exposed or buried in the ground or the swamp.

The negative return line from the vessel, approximately 600 feet long, was number 4/3 rubber covered cable used as a single conductor.

The impregnated graphite anodes were special 4-inch-diameter by 40-inches long, intended for test purposes only. The bed of ten anodes was installed on April 20, 1950, in the swamp between the rectifier and the vessel, approximately 450 feet directly in shore of the bow. Each anode was buried substantially horizontal and lengthwise to the longitudinal axis of the vessel, in the very watery areas of the swamp about 1 foot beneath the surface, pushed into place by hand, there being no need for digging holes. The ten anodes were all placed within a 60-foot-wide area (equal to vessel breadth), spaced from 3 to 10 feet apart.

The swamp water resistance at this site was calculated to be 6,000 ohm-centimeters plus or minus 20 per cent, depending upon tide conditions. Typical analyses of water samples, taken March 20, 1950, are as follows:

	pH	Sodium Chloride, Grains per Gallon
Surface	6.55	0.89
Bottom	6.55	0.94

The submerged surface of the vessel was calculated to be approximately 20,000 square feet. Copper sulphate half-cell readings taken before current was applied showed the hull of the vessel to be −0.59 volt, indicating a severe corroding condition.

The anode installation was energized at 60 volts at 11:00 a.m. on April 20, 1950. A remote potential reading, with current on, at 11:20 a.m. showed negative 1.36 volts between the vessel and the copper sulphate half cell, while a similar reading taken at the same time but approximately 40 feet from the stern showed negative 1.24 volts at 6 inches beneath the surface and negative 1.32 volts at a depth of 10 feet.

Remote readings were again taken at 4:00 p.m. showing negative 1.6 volts while those taken close to the stern were negative 1.5 volts at 6 inches below the surface and negative 1.54 volts at a depth of 10 feet. The bed was discharging 8 amperes from 11:00 a.m. for the rest of the day and overnight, April 20–21, 1950. Official test readings started at 8:30 a.m., April 21, 1950.

Changes in salinity caused erratic readings with current on, whereas the more significant readings, taken with the current off, were relatively stable. Readings taken close to the bow were generally lower than those taken near the stern, probably due to the bow receiving less current because of direct contact with mud in the shallow water.

Conclusions

In consideration of the general acceptance of −0.78-volt half-cell reading as the theoretical level at which protection is obtained, the protection indicated by the readings was achieved with the minimum amount of current. Complete polarization was reached very slowly, but maintained at a consistently high level even though there were varying water resistances the effect of which is more pronounced in test number 3. Remote half-cell readings with current on are not included in the data in Figure 2(B).

On June 5, 1950, four more 4 by 40-inch anodes and two 3 by 60-inch anodes were added to the ten already in service. On January 2, 1952, four more 3 by 60-inch anodes were added, bringing the total to 20.

Test Number 3

The third test was conducted at another fleet site. The vessel selected for the test was the nearest to shore of a series of ships anchored side by side, and was separated from the others by 20-foot fenders. Its submerged surface is approximately 20,500 square feet.

The anodes consisted of eleven 3 by 60-inch impregnated graphite, spaced about 12 feet apart along a feeder cable, laid broadside to the test vessel, about 250 feet away in shallow water having an average depth of 5 feet.

Potential readings on the hull were taken with a high-resistance voltmeter (internal resistance 200,000 ohms per volt) and a copper sulphate half cell. The remote reading before current was applied was −0.465 volt. The data taken are given in Table III.

The rectifier was out of service and the test was interrupted for about 45 days. Just before again energizing the test on February 8, 1951, average remote was 0.405 volt. On September 23, 1951, the current was off for 12 hours.

Conclusions

Polarization was comparatively rapid and a high level was achieved although some extreme resistance changes were encountered because of changing salinity. Since the periods of low resistance were relatively short, the polarization level was not greatly affected. The data for test number 3 are summarized in Figure 2(C).

No Discussion

Synthesis of Feedback Control System by Phase-Angle Loci

YAOHAN CHU
STUDENT MEMBER AIEE

THE synthesis of a feedback control system by means of phase-angle loci originates from Evan's root-locus method[1] which enables one to find all possible roots of the characteristic equation of a feedback control system. Once the roots are chosen, the transient response then may be worked out. Thus, it allows the synthesis of feedback control system in time domain. However, the plotting of the root locus requires some sort of guesswork. This is not surprising because the analytical difficulty in extracting the roots of the characteristic equation is imposed on the graphical work, yet help in the effort to solve the problem can be given.

In this paper, the phase-angle loci are introduced, by means of which not only the plotting of the root locus can be carried out in a systematic manner and the guesswork thus is eliminated but also the root locus can be reshaped to a desirable one. Furthermore, synthesis for a widely accepted criterion of transient response is enumerated, and approximate equations are given so that the transient response may be evaluated without actually working out the transient solution. This is an important link in the synthesis by this method, because not only synthesis may be quickly carried out then but also certain relations between the poles and zeros of the closed-loop transfer function and the transient response are formulated quantitatively then. A brief account of the root-locus method is given in the Appendix.

Phase-Angle Loci

For a given transfer function $G(s)$ the plot on the complex s-plane of the family of the following relation

$$\text{Ang}\, G(s) = \phi \tag{1}$$

for various values of the phase-angle ϕ is called the phase-angle loci, where s $(=\sigma+j\omega)$ is the complex variable. Note that the particular phase-angle locus with ϕ equal to ±180 degrees is the root locus.

Two basic forms of the phase-angle loci are introduced now. They are called basic because of their simple geometric forms. One family is of straight lines, and the other of circles. The first basic form is for the transfer function as given in equation 2 (The power of equation 2 may be values other than ±1, which also will give a family of straight radial lines. The value ±1 is used here because of simplicity in discussion.)

$$G(s) = (s+s_1)^{\pm 1} \tag{2}$$

where either positive or negative sign is taken and s_1 can be zero, a real, or complex number. It is a transfer function of one pole (for negative sign) or one zero (for positive sign) at any place in the s-plane. This phase-angle locus is a family of straight lines. This can be shown by applying equation 2 to equation 1; thus

$$\text{Ang}\,(s+s_1)^{\pm 1} = \phi \tag{3}$$

By substituting $s=\sigma+j\omega$ and $s_1=\sigma_1+j\omega_1$ into equation 3, the result is

$$\frac{\omega+\omega_1}{\sigma+\sigma_1} = \tan(\pm\phi) \tag{4}$$

This is a straight line for each value of ϕ passing through the point (σ_1, ω_1) and with a slope of $\tan(\pm\phi)$ on the s-plane. Therefore, it is a family of straight lines for various values of ϕ. Figure 1 shows such a family of straight lines for various values of ϕ with s_1 equal to unity. It is obvious that the construction of such a family is very simple.

The second basic form is for the transfer function of the form

$$G(s) = \frac{s+\sigma_1}{s+\sigma_2} \tag{5}$$

where σ_1 and σ_2 are real numbers. It is a transfer function of one pole and one zero on the real axis of the s-plane. It can be shown by similar analytic derivation that each phase-angle locus is a circle with center at the point $[-(\sigma_1+\sigma_2)/2, r\cos\phi]$ where the radius r is

$$r = \left|\frac{\sigma_2-\sigma_1}{2\sin\phi}\right| \tag{6}$$

Figure 2 shows such phase-angle loci for the case $\sigma_2 > \sigma_1$ (or pole at $-\sigma_2$ and zero at $-\sigma_1$). This family is also very simple to construct. The transfer function, equation 5, represents an integral or a derivative function of a simple passive network. Figure 2 shows the case where it represents a derivative network.

Although it is possible to plot the gain loci in an analogous fashion to the phase-angle loci, yet such gain loci do not contribute much usefulness to the topics in this paper. Hence, no further discussion is attempted.

Hydro-Analogy (or Electrostatic Field Analogy)

The magnitude and angle functions of the transfer function $G(s)$ are two conjugate functions (that is, the phase-angle loci are orthogonal to the gain loci). The phase-angle loci are therefore nothing more than a pattern of source-and-sink flow, where the loci are the stream lines (or a pattern of electrostatic field, where the loci are the electric lines of force). The poles of $G(s)$ for convenience may be considered as sources and zeros as sinks (or poles as positive charges and zeros as negative charges). Figure 1 shows the flow pattern of a single source, with the radial lines as the stream lines. Figure 2 is a pattern of one source and one sink flow with circles as the stream lines. Figure 3 shows the flow pattern of two sources. These patterns are invariant with respect to the actual location of poles and zeros on the s-plane, although the significance in connection with the transient response will be different. If there are only simple poles and zeros, then all sources and sinks (or charges) are of the same strength. By means of analogy, the pattern of the phase-angle loci for a given configuration of poles and zeros of the transfer function $G(s)$ may be visualized before actually being drawn. Consequently, this analogy is helpful in constructing the phase-angle loci.

It now may be summarized that the properties of the phase-angle loci are

a. they are symmetrical with respect to the real axis;

b. the pattern of the phase-angle loci depends on the configuration of the poles and zeros of $G(s)$ but is independent of the axes of the s-plane; and

c. each phase-angle locus will approach the asymptote with the slope

---

Paper 52-297, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted June 24, 1952; made available for printing September 22, 1952.

YAOHAN CHU is with Massachusetts Institute of Technology, Cambridge 39, Mass.

The author wishes to express his appreciation and thanks to Prof. Walter McKay for his kind encouragement and guidance and also to Prof. Donald P. Campbell for his valuable comments and for the opportunity to present this material in the Servomechanisms course.

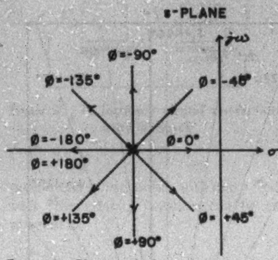


Figure 1. Phase-angle loci of the first basic form

$$\tan\alpha = \tan\left(\frac{\pm\phi}{n-m}\right) \qquad \text{for } n \neq m \qquad (7)$$

and the asymptote will pass through the point, known as centroid, on the real axis at a distance L from the left of the origin

$$L = \frac{\sum_{i=1}^{n} p_i - \sum_{j=1}^{m} z_j}{n-m} \qquad (8)$$

where n is the number of poles, m the number of zeros, p_i the values of the poles, and z_j the values of zeros of the transfer function $G(s)$. The second property enables a family of phase-angle loci to have several uses sometimes.

Superposition Method

The phase-angle loci of any given transfer function can be constructed by the superposition method. This method is first to split a transfer function into several transfer functions, so that the phase-angle loci of each have a basic form. Then by superimposing two of these several splitted transfer functions at a time, the required phase-angle loci finally will be obtained. The amount of work will depend on the complexity of the given transfer function. It is clearer to explain this method by an illustration.

Suppose the given transfer function is

$$G(s) = \frac{K}{s(s+1)} \qquad (9)$$

By definition, the phase-angle loci of this transfer function are

$$\text{Ang}\,\frac{1}{s} + \text{Ang}\,\frac{1}{s+1} = \phi$$

The superposition method first splits this expression into the following two expressions

$$\text{Ang}\,\frac{1}{s} = \phi_1 \quad \text{and} \quad \text{Ang}\,\frac{1}{s+1} = \phi_2$$

The phase-angle loci of the first expression are a family of radial lines similar to Figure 1 except that the pole is at origin, and those of the second one are of the same pattern except that the pole is at the point $(-1, 0)$. Now these two families are superimposed on the same s-plane that is shown in Figure 4. The phase-angle locus of $G(s)$ for any value ϕ is the curve of all the points of intersection of these two families of lines, at which the sum ϕ_1 and ϕ_2 is equal to ϕ. For example, point a in Figure 4 is the point of intersection of the phase-angle locus of ϕ_1 equal to -120 degrees and that of ϕ_2 equal to -60 degrees. There, the sum of ϕ_1 and ϕ_2 is -180 degrees. Similarly, points b, c, and d are located. The curve passes the points a, b, c, and d is the phase-angle locus with ϕ equal to -180 degrees (or the root locus). The phase-angle locus with ϕ equal to other values can be constructed similarly. The phase-angle loci of the given $G(s)$ are the pattern as shown in Figure 3, and also are shown in Figure 5 where only those loci on the upper-half plane are drawn.

As a second illustration, the following transfer function is selected

$$G(s) = \frac{K}{s(s+1)(s+5)} \qquad (10)$$

By definition, the phase-angle loci are

$$\text{Ang}\,\frac{1}{s(s+1)} + \text{Ang}\,\frac{1}{s+5} = \phi$$

The method of superposition first splits this expression into the following two expressions

$$\text{Ang}\,\frac{1}{s(s+1)} = \phi_1 \quad \text{and} \quad \text{Ang}\,\frac{1}{s+5} = \phi_2$$

The phase-angle loci of the first expression are shown in Figure 5 and those of the second expression are similar to those shown in Figure 1 except that the pole now is at the point $(-5, 0)$. If the second phase-angle locus is superimposed on the first one, then the phase-angle locus with a certain value of ϕ is obtained from the points of intersection of these two families, at which the sum ϕ_1 and ϕ_2 is equal to ϕ, the same procedure as the previous example. The phase-angle loci of this second example are shown in Figure 6, where only several loci on the upper-half plane are drawn. The phase-angle loci in both Figures 5 and 6 fulfill, of course, equations 7 and 8. Their construction will be facilitated by using these two equations.

Figure 6 shows the new pattern of Figure 5, which is effected by an additional simple lag of $1/(s+5)$. The stream lines of Figure 5 are being pushed toward the right, as shown in Figure 6. If additional poles are introduced at the left of the original poles, the stream lines again will be pushed further toward the right. The extent of the effect depends on the actual location of the source and its strength. The root locus (which is the phase-angle locus of ±180 degrees) in general will bend toward the right of the imaginary axis of the s-plane because practical systems always consist of lags and thus provide the possibility of instability.

Reshaping of the Root Locus

The superposition method may be used not only to construct the phase-angle loci

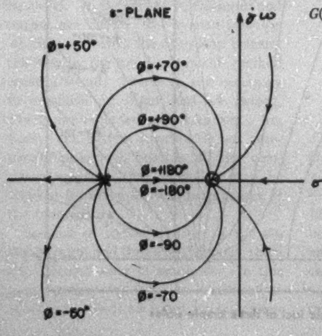


Figure 2 (left). Phase-angle loci of the second basic form

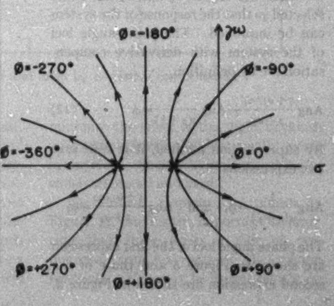
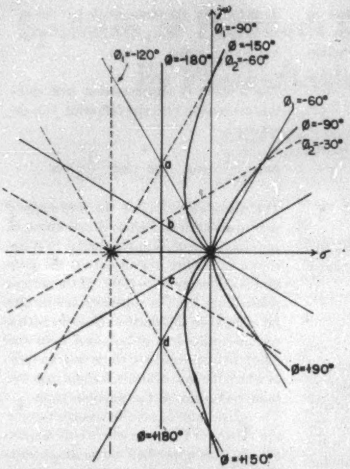


Figure 3 (right). Phase-angle loci of the transfer function with two poles



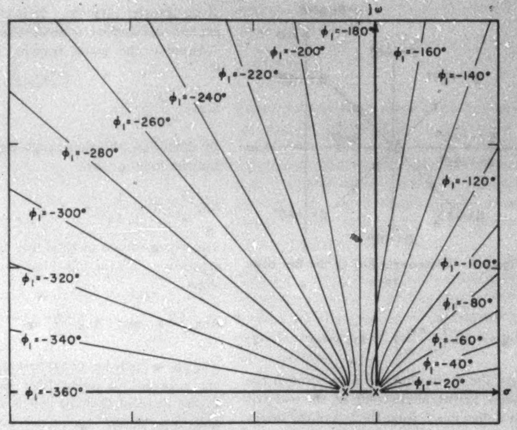


Figure 4 (left). Construction of phase-angle loci of $G(s)=\frac{K}{S(S+1)}$

Figure 5 (above). Phase-angle loci of two simple poles

of a given transfer function but also to reshape the original root locus. An example is employed to illustrate the procedure. Figure 7 shows a simple feedback control system with the transfer function of the proportional controller $G_1(s)$

$$G_1(s)=K_1$$

and the transfer function of the controlled member $G_2(s)$

$$G_2(s)=\frac{K_2}{s(s+1)}$$

For convenience, let $G(s)=G_1(s)\ G_2(s)$. The phase-angle loci of $G(s)$ are those shown in Figure 5, where the locus with ϕ equal to ±180 degrees is the root locus. If a derivative network is to be incorporated in the controller, its transfer function becomes

$$G_1(s)=K_1\frac{(s+\sigma_l/\alpha_l)}{(s+\sigma_l)} \tag{11}$$

in which the values σ_l and α_l are to be selected so that the response of the system can be improved. The phase-angle loci of the system with derivative compensation are by definition

$$\text{Ang}\,\frac{s+\sigma_l/\alpha_l}{s+\sigma_l}+\text{Ang}\,\frac{1}{s(s+1)}=\phi \tag{12}$$

By superposition method, it is split into two expressions

$$\text{Ang}\,\frac{1}{s(s+1)}=\phi_1 \quad\text{and}\quad \text{Ang}\,\frac{s+\sigma_l/\alpha_l}{s+\sigma_l}=\phi_2$$

The phase-angle loci of the first expression are shown in Figure 5 and those of the second expression are shown in Figure 2. Superimpose the second expression on the first as shown in Figure 8. The new root locus can similarly be found as shown in the preceding section. The comparison between this new root locus with that of Figure 5 shows the change of the original root locus by the derivative controller. If the second one with the pole-to-zero distance constant is superimposed at the different locations along the real axis of the first one as shown in Figures 9, 10, and 11, various reshaped root loci are obtained. It is not necessary when reshaping to keep constant the pole-to-zero distance in the phase-angle loci of Figure 2. It would be advantageous to keep the value α_l constant instead. The magnitude equation of the system now becomes

$$K=\left|\frac{s(s+1)(s+\sigma_l)}{(s+\sigma_l/\alpha_l)}\right| \tag{13}$$

Since this example may be regarded as a

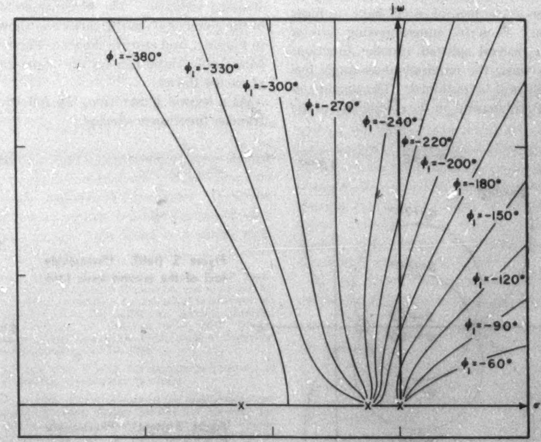


Figure 6. Phase-angle loci of three simple poles

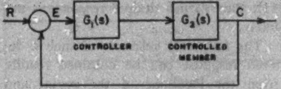


Figure 7. A feedback control system with unity feedback

position servomechanism with unity feedback, the velocity constant can be expressed[2] as

$$K_v=K/\alpha_1$$

Equations 13 and 14 will be used later to calculate the velocity constant.

Synthesis in the Time Domain

In general, synthesis in the time domain is more difficult than that in the frequency domain. However, one often desires to know the transient response, if one designs by the frequency response method, especially if specification of a system is on the transient basis. Various methods are available to obtain the transient response from the open-loop or closed-loop frequency response, such as Floyd's method,[2] Chestnut and Mayer's charts,[3] the Harris-Kirby-von Arx method,[4] and Biernson's method.[5] In this paper, an attempt is made to synthesize the feedback control system directly in the time domain by the introduction of phase-angle loci to Evan's root-locus method, and also to develop an important link in the synthesis procedure. This is a convenient but quite accurate relation between the chosen roots from the root locus and the transient response, so the transient response may be evaluated quickly without actually working out the transient solution.

In the analysis of a system, the input and the system function (or the closed-loop transfer function) are known; the output or the response of the system must be found. In the synthesis, the input and output are known; the system function which is related to the open-loop transfer function by the root locus in this method must be found. Therefore, it is necessary to establish an input and an output, which are regarded to be known. Actually, the system function is not completely unknown, because all the components except the controller may be selected from the known performance of the components.

Because of various choices of the accepted input and the desired output, there are different ways of synthesis. In this paper, a unit step function is selected as the input. Such an input is convenient and widely used. For the desired output, a response with a slight oscillation and slight overshoot, as shown in Figure 12, is selected. The actual desired number of oscillations and actual desired amount of overshoot will be left to be decided in each particular application. Such a slight underdamped response will give a somehow faster response and also will minimize the effect of the slight backlash or coulomb friction which usually exists in a system. Such a desired output is also widely accepted and used.

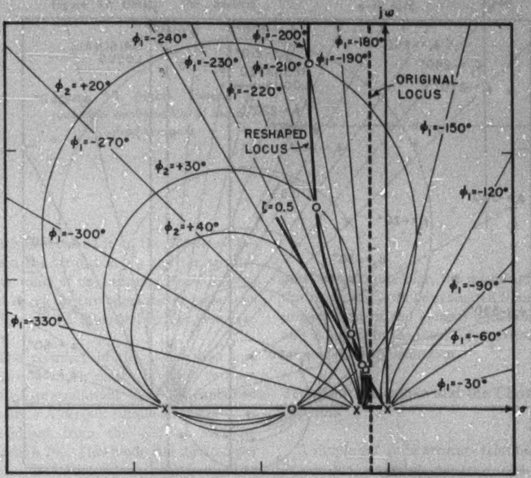


Figure 8. Reshaping of the root locus by derivative compensation

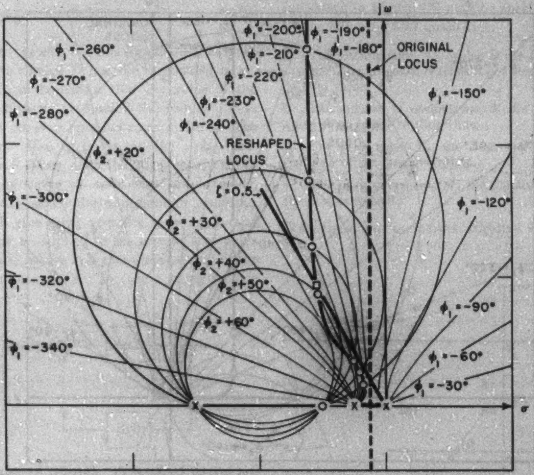


Figure 9. Reshaping of the root locus by derivative compensation

The desired response, as shown in Figure 12, implies the important assumption that there exists a pair of predominate complex roots in the system function. This may be shown clearly on

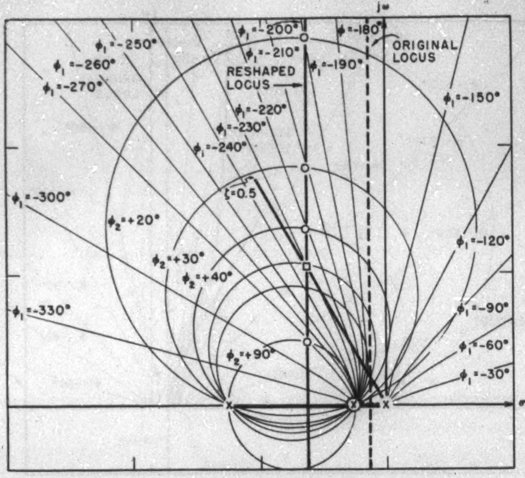


Figure 10. Reshaping of the root locus by derivative compensation

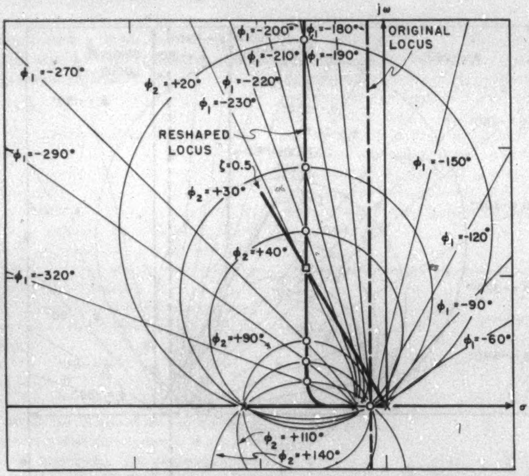


Figure 11. Reshaping of the root locus by derivative compensation

the *s*-plane. Figure 13 shows such a plane, where the pair of complex roots is the predominate pair. Other roots may exist at the left of this complex pair. Other pairs of complex roots also may exist if they occur relatively far away from the predominate pair; they are not shown in Figure 13. It is important to note there that the pair of the predominate complex roots is the nearest to the imaginary axis from the left than any other real or complex roots. Thus, the conclusion now is reached that the roots should be so chosen from the root locus that their configuration is of the pattern shown in Figure 13. If the root locus does not offer any possibility of such a pattern (as shown in the case of Figure 11), then such a desired response is unlikely.

The frequency response of such a desired response can be obtained readily from the locations of the roots (and zeros) of the system function by drawing the log-modulus versus log-frequency plot from these roots (and zeros). Thus, a correlation exists between the desired transient response just selected and the frequency response. Since there is a pair of predominate complex roots in the system function, the frequency response will show a peak in its amplitude plot. Such a frequency response (where the amount of peak and the frequency at which the peak occurs are the two most important results) is a widely accepted one. Consequently, the desired transient response just selected agrees with the criterion used in the frequency response.

The Approximate Equations

The desired output which has been established in the last section may be represented in the *s*-plane by the configuration of roots and zeros of the system function as shown in Figure 13. The next step in the process of establishing the link is to relate quantitatively this configuration with the transient response. This relation is the general transient solution. Because of the complexity of this general solution, approximate relations will be formulated in this section, and their significance, as well as the illustrations, will be explained in the next section.

For a given feedback control system, Figure 14, the system function C/R is

$$\frac{C}{R}=\frac{G(s)}{1+G(s)H(s)}=K\frac{A(s)}{B(s)} \tag{14}$$

where

$$A(s)=\prod_{j=1}^{m}(s-z_j)$$

$$B(s)=\prod_{k=1}^{n}(s-q_k)$$

In this expression it is assumed that no multiple roots exist and the order of $A(s)$ is smaller than that of $B(s)$. These two assumptions are true in most applications. The values z_j and q_k may either be real or complex. By using Laplace inverse transformation, the transient solution of equation 14 for unit step input with all zero initial conditions may be expressed as

$$c(t)=K\frac{A(0)}{B(0)}+\sum_{k=1}^{n}\frac{KA(q_k)}{q_kB(q_k)}\epsilon^{q_kt} \tag{15}$$

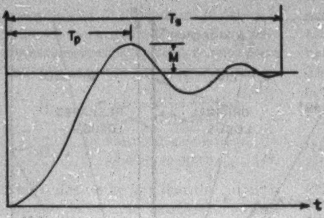


Figure 12 (left). The desired response

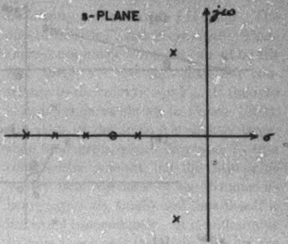


Figure 13 (right). Pattern of root-zero configuration of the desired response

If complex roots are combined, it may be written as

$$c(t)=K\frac{A(0)}{B(0)}+\sum_{q_k \text{ real}}\frac{KA(q_k)}{q_kB'(q_k)}\epsilon^{q_kt}+$$

$$\sum_{q_k=\sigma_k+j\omega_k}2\left|\frac{KA(q_k)}{q_kB'(q_k)}\right|\epsilon^{\sigma_kt}$$

$$\cos\{\omega_kt+\text{Ang }A(q_k)-\text{Ang }q_k-\text{Ang }B'(q_k)\} \quad (16)$$

This is the general transient solution within the limits of the assumptions. The output response is the summation of all simple modes due to simple roots and of all complex modes due to complex roots in addition to a constant term.

The desired output response, shown in Figure 12, is virtually assured by choosing the type of root configuration shown in Figure 13. An assumption is now made in order to reach a simple but practical result: It is assumed that all the modes of equation 16 may be neglected except the first constant term and except those due to the pair of predominate roots. It is important to realize that the assumption here does not neglect all other roots but does neglect all other modes because these other modes die out quickly. With such an assumption, equation 16 becomes approximately

$$c(t)=\frac{KA(0)}{B(0)}+2\left|\frac{KA(s_0)}{s_0B'(s_0)}\right|\epsilon^{-\sigma_0t}$$

$$\cos\{\omega_0t+\text{Ang }A(s_0)-\text{Ang }(s_0)-\text{Ang }B'(s_0)\} \quad (17)$$

where $s_0=-\sigma_0+j\omega_0$ is the predominate roots. By differentiating the output with respect to time and equating the result to zero, the time to reach the peak of the first overshoot, to be called T_p is

$$T_p\text{(peak time)}=\frac{1}{\omega_0}\left\{\frac{\pi}{2}-\text{Ang }A(s_0)+\text{Ang }B'(s_0)\right\} \quad (18)$$

The amount of the first overshoot, M, is

$$M\text{ (peak overshoot)}=\frac{2\omega_0}{\sigma_0+\omega_0}\left|\frac{KA(s_0)}{B'(s_0)}\right|\epsilon^{-\sigma_0T_p} \quad (19)$$

which is obtained from the difference between the output at T_p and the steady-state value of the output. Equations 18 and 19 are the two approximate equations which will be discussed in the next section.

The accuracy of the assumption just discussed which results in these two approximate equations may be explained readily. A certain simple mode $D\epsilon^{-\sigma_1t}$ is considered from the general solution, equation 16. This mode will die to 5 per cent of its initial value when t is equal to $3/\sigma_1$. If T_p is larger than this value, that is

$$T_p\geq\frac{3}{\sigma_1} \quad \text{or} \quad \sigma_1\geq\frac{3}{T_p} \quad (20)$$

then the effect of this simple mode is small even though D may be reasonably large. (It can be shown that the value of the coefficient D is small if this root $s=-\sigma_1$ is not too close to other roots and if no zero very close to origin on the s-plane exists.) Thus, equation 20 shows the minimum value of σ_1 of any simple mode, above which the approximate equations will give quite accurate results. It also may be applied to the case of a complex mode, where σ_1 is then the real part of the complex roots. In case a certain mode must be accounted for, the magnitude of this mode at T_p can be evaluated from the proper term of the general equation 16 and is added to the result of equation 19 as a second approximation.

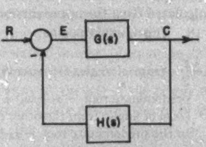


Figure 14 (above). A feedback control system

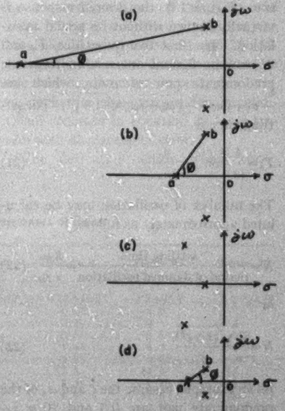


Figure 15 (right). The effect on T_p and M by the location of a third root to a pair of complex roots of a system function on s-plane

(a) ao/ab ≐ 1 and φ is small
(b) ao/ab slightly greater than one and 0 degree < φ < 90 degrees
(c) root configuration which does not give the desired response
(d) ao/ab ≫ 1 and 0 degree < φ < 90 degrees

Transient Response and the Chosen Roots

A simple yet quite accurate relation between the transient response and the chosen roots from the root locus is the last but most important link in the synthesis procedure. The desired response (as shown in Figure 12) may be specified by the following four quantities:

1. The settling time, T_s, which is the time at which the response reaches 2 per cent of its final value.
2. The number of oscillations, N, for the interval up to the settling time.
3. The peak time, T_p, the time to reach the peak of the first overshoot.
4. The peak overshoot, M, the amount of the first overshoot.

With these four quantities, together with

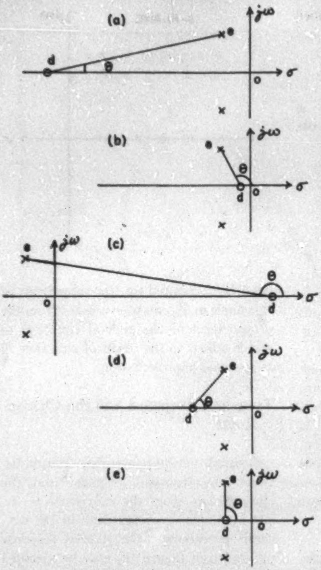


Figure 16. The effect on T_p and M by location of a zero of a system function on s-plane

(a) de/do ≐ 1 and θ is small
(b) de/do ≫ 1 and 90 degrees < θ < 180 degrees
(c) de/do ≐ 1 and θ is near 180 degrees
(d) de/do slightly greater than one and θ ≐ 60 degrees
(e) de/do reasonably greater than one and θ ≐ 90 degrees

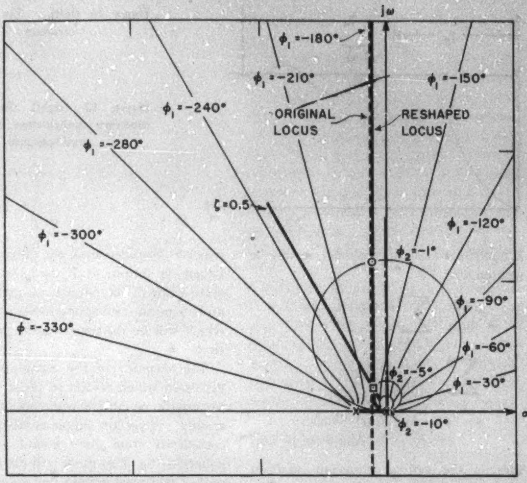


Figure 17 (above). Reshaping of root locus by integral compensation

the pattern of the chosen root-configuration (Figure 13), the desired response is virtually known without its actual calculation. The first two quantities, T_s and N, can be figured out readily from the predominate pair of roots, which are $-\sigma_0+j\omega_0=-\zeta\omega_n+j\omega_n\sqrt{(1-\zeta^2)}$. The settling time is

$$T_s=\frac{4}{\sigma_0} \quad \text{or} \quad =\frac{4}{\zeta\omega_n} \tag{21}$$

The number of oscillation may be calculated approximately as follows

$$N=\frac{\text{settling time}}{\text{period of damped oscillation}}=\frac{2}{\pi}\frac{\omega_0}{\sigma_0} \tag{22}$$

or

$$N=\frac{2}{\pi}\frac{\sqrt{(1-\zeta^2)}}{\zeta} \tag{23}$$

As an example, assume the ζ and ω_n of the predominate pair are 0.5 and 20 π respectively. Then

$$T_s=\frac{4}{\zeta\omega_n}=\frac{4}{0.5\times 20}=0.127 \text{ second}$$

and $N=1.11$ (that is, one overshoot and one undershoot). That the settling time and the number of oscillation are determined by predominate roots only is a result of the assumption of the existence of the pair of the predominate roots so that all other roots die out earlier than the predominate pair.

The third quantity, T_p, the time to reach the peak of the first overshoot, is to be calculated from the approximate equation 18. This is explained as

$$\omega_0 T_p=\frac{\pi}{2}-(\text{sum of angles from zeros to the predominate root } -\sigma_0+j\omega_0)+(\text{sum of angles from the other roots to the predominate root } -\sigma_0+j\omega_0) \tag{24}$$

Thus, it may be concluded that zeros decrease T_p and additional roots increase T_p. If the system is known to be approximately of the second order and without any zero, then the peak from equation 24 is

$$T_p=\frac{\pi}{\omega_0} \tag{25}$$

Because of the relation $\omega_0=\omega_n\sqrt{(1-\zeta^2)}$, another conclusion from equation 24 is that the peak time is inversely proportional to the undamped natural frequency ω_n of the predominate roots. The higher the ω_n is, the smaller the T_p is. This is a well-known result, yet the truth of this statement is shown now and is limited to the assumptions thus made.

The last quantity, M, the amount of the first overshoot, is to be calculated from the approximate equation 19. It is the product of the following two terms

(a) $\epsilon^{-\sigma_0 T_p}$ (this quantity is always less than unity as σ_0 and T_p are positive) (26)

$$\text{(b)} \quad \frac{K}{\omega_n^2}\times\frac{\text{(product of distances from zeros to the predominate root } -\sigma_0+j\omega_0)}{\text{(product of distances from all roots to the predominate root } -\sigma+j\omega_0 \text{ excluding the distance between the two predominate roots)}} \tag{27}$$

It is obvious from the first term (a) that the smaller the T_p the larger the M; thus the choice of T_p and M requires a compromise. It also may be examined from the equation

$$\sigma_0 T_p=\frac{\sigma_0}{\omega_0}\left\{\begin{matrix}\text{same terms as correspond-}\\ \text{ing ones of equation 24}\end{matrix}\right\} \tag{28}$$

M is thus dependent of the ratio σ_0/ω_0 which is equal to $\zeta/\sqrt{(1-\zeta^2)}$. Then, when the damping ratio ζ is small, M is large, a conclusion consistent with the known result. However, a quantitative relation as well as its limitation are shown now.

The effect of the second term (b) on the magnitude of M depends on K, which is known for an actual problem. For a zero-displacement error system, K must have the following value (from equation 14)

$$K=\frac{\prod_{k=1}^{n} q_k}{\prod_{j=1}^{m} z_j}=\frac{\text{(product of distances from all roots to the origin)}}{\text{(product of distances from all zeros to origin)}} \qquad (29)$$

This is due to the fact that the output is equal to the input at steady state for step input in such a system. By substituting equations 28 and 29 into 19, the result is

$$M=\frac{\text{(product of distances from all roots to the origin excluding distances from the two predominate roots to origin)}}{\text{(product of distances from all roots to the predominate root } -\sigma_0+j\omega\text{, excluding the distance between the predominate roots)}}$$
$$\times\frac{\text{(product of distance from zeros to the predominate root } -\sigma_0+j\omega_0\text{)}}{\text{(product of distance from all zeros to origin)}}\times\epsilon^{-\sigma_0 T_p} \qquad (30)$$

If the angles θ and ϕ represent respectively the angles which the zero and the additional root make with the predominate root $(-\sigma_0+j\omega_0)$ (as shown in Figures 15 and 16), then equation 24 may be expressed as

$$T_p \doteq \frac{1}{\omega_0}\left\{\pi-\sum_{\text{all zeros}}\theta+\sum_{\substack{\text{all roots except}\\ \text{the predominate pair}}}\phi\right\} \qquad (31)$$

Furthermore, if the lengths *ao* and *ab* represent respectively the distances from the additional root to the origin and to the predominate root $(-\sigma_0+j\omega_0)$ as shown in Figure 15 and *do* and *eo* represent respectively the distances from the zero to the origin and to the predominate root $(-\sigma_0+j\omega_0)$ as shown in Figure 16, then equation 30 may be expressed as

$$M=\underbrace{\left\{\prod_{\substack{\text{all roots except}\\ \text{the predominate pair}}}\frac{ao}{ab}\epsilon^{-\frac{\sigma_0}{\omega_0}\phi}\right\}}_{\text{due to additional roots}}$$
$$\underbrace{\left\{\prod_{\text{all zeros}}\frac{de}{do}\epsilon^{+\frac{\sigma_0}{\omega_0}\theta}\right\}}_{\text{due to zeros}}\underbrace{\left\{\epsilon^{-\frac{\sigma_0}{\omega_0}\pi}\right\}}_{\text{constant}} \qquad (32)$$

Thus, it is shown in equation 32 that the first and second brackets indicate respectively the effect on M (amount of first peak overshoot) due to the additional roots and to all zeros. The third bracket is regarded as constant for a given pair of complex roots. Note that these effects depend on the relative location among the roots, the zeros, and the origin. Also note from equation 31 that T_p (the peak time) does not depend on these distances but depends only on the angles ϕ and θ. The effects of the first and second brackets are further discussed in the following two paragraphs.

The first bracket of equation 32 gives the effect on M due to the additional roots other than the predominate pair of complex roots. If a root is far away at the left of the complex roots as shown in Figure 15(A), the ratio *ao/ab* is near unity and the angle ϕ is small. The effects on M and T_p are both small. If the root is at the location shown in Figure 15 (B), the ratio *ao/ab* is slightly larger than unity but the angle ϕ is of certain magnitude between 0 degrees and 90 degrees. The value of the first bracket of equation 32 will be smaller than that of Figure 15 (A); thus M is somewhat reduced, but T_p is somewhat increased as the angle ϕ is not small. If the root is located at the position shown in Figure 15(C) in which the assumption of the existence of a predominate pair of complex roots fails, the response will have a "long tail" which is regarded as undesirable. In addition, the approximate equations do not hold. Therefore, this case is disregarded. If the pair of complex roots and the real root are located as shown in Figure 15(D), the value of ϕ is nearly the same as that of Figure 15(B), so its effect on T_p is also near that of Figure 15(B), but the value of the ratio *ao/ab* is very much greater than unity, which results in a large value of M, or the amount of first peak overshoot. It may be deducted further that the response will be highly oscillatory if the three roots are coinciding, that is, a triple root. Since highly oscillatory response is undesirable, the root configuration such as that of Figure 15(D) should be avoided. If there is more than one additional root, their effects are combined in a manner as shown by equations 31 and 32.

The second bracket of equation 32 gives the effect on M due to the zeros. If a zero is far away at the left of the origin and of the predominate roots as shown in Figure 16(A), the ratio of the distances *de/do* is near unity and the angle θ is small. The effects on both T_p and M are small. Thus, there is not much overshoot caused by the zero there, yet the zero does not contribute to the reduction of T_p. If the zero is near origin as shown in Figure 16(B), the value of the angle θ is slightly larger than 90 degrees, which decreases T_p to a considerable amount, but the ratio *de/do* is very large (because of the distance *do* being extremely small) and thus the M is increased tremendously. It is undesirable to have a zero near the origin. If the zero is at the right half of the *s*-plane as shown in Figure 16(C), the ratio *de/do* is near unity, but the angle θ is between 90 degrees and 180 degrees. T_p will be reduced but M will be increased to a certain extent. The value of M may be acceptable if the ratio σ_0/ω_0 is chosen to be small. This case gives more desirable results than that of Figure 16(B), but a zero at the right-half plane may be difficult to be realized. If the zero is located slightly to the left of the complex roots, the ratio *de/do* is slightly larger than unity, but the angle θ is still quite appreciable so that T_p is somewhat reduced without an excessive large value of M. In ordinary applications where the reduction of T_p is desirable but an excessive increase of M is undesirable, the location of the zero in Figure 16(D) is a good choice of compromise between T_p and M. If slightly larger M is tolerable; the zero location that is shown in Figure 16(E) is also a good choice. If there is more than one zero, their combined effects are related in a manner shown by equations 31 and 32.

In short, the discussions in the two preceding paragraphs give a qualitative visualization of the effects on M and T_p by the configuration of the roots and zeros. The quantitative evaluation can be achieved by equations 31 and 32. Although these discussions are carried out on the zero-displacement error system, the result is the same except for an additional constant in equation 29 for a system with displacement error.

The derivative action which corre-

Table I. Transient Responses of Roots

Term	Symbol	Original	Figure 8	Figure 9	Figure 10
Complex roots	s_0	$-0.5+j0.87$	$-0.75+j1.25$	$-2.2+j3.7$	$-2.5+j4.25$
Real roots	σ_1		-6.55	-4.0	
	αl		2.33	3	5
Damping ratio	ζ	0.5	0.5	0.5	0.5
Undamped natural frequency	ω_n	1.0	1.45	4.3	4.9
Peak time, seconds	T_p	3.6	2.25	0.62	0.74
Peak overshoot, per cent	M	16.5	17.5	14.8	15.7
Settling time, seconds	T_s	8.0	5.32	1.82	1.6
Gain	K	1.0	4.6	11.9	18
Velocity constant	K_v	1.0	1.97	3.95	3.6

sponds to zeros usually is employed for either speeding up the response or stabilizing the system. The previous discussion about zero location is for speeding up the response. For stabilizing the system, the reshaped root locus should be pushed toward the left of the imaginary axis of the s-plane so that larger values of the gain of the system are possible. This was not shown in the previous example (Figures 8 to 11) because the original system is of the second order, which is stable for any value of the gain.

As illustrations, the transient responses of the roots chosen from the root loci of Figures 8 to 10 and the response chosen from the original root locus of Figure 5 are given in Table I.

All the roots in Table I are so chosen that the predominate roots have a damping ratio of 0.5. N may be calculated from equation 23 and is equal to 1.1 (one overshoot and one undershoot) for all these cases. It is noted that no root-zero configuration similar to that of Figure 13 is possible from the root locus of Figure 11, thus no desired response such as that of Figure 12 is obtainable. The preceding equations do not apply, and no result is calculated from Figure 11. The zero in Figure 10 just cancels out one pole, and the result is a simple second-order system. It seems from the comparison of the results in Table I that the roots chosen from Figure 9 or 10 are desirable, because considerable increase in undamped natural frequency (also considerable decrease in T_p) results with a reasonable amount of the first overshoot, in addition to a certain increase in velocity constant. It is suggested to reshape with larger values of α_i with zero near the location of either Figure 9 or 10 to further improve the results. Furthermore, a slightly different damping ratio may be used if so desired.

Integral Compensation

The reshaping of the root locus, described in a previous section, is done by means of a derivative function for the purpose of speeding up the response. This is known as "derivative compensation." Many cases arise where it is desired to increase the gain of the system to improve its steady-state performance, and this is achieved by "integral compensation." With integral compensation, it is desirable to keep the original dynamic performance. This means that the root locus will be changed a little by the introduction of the integral compensation, particularly at those portions of the root locus where the roots have been chosen.

The example in the section entitled "Reshaping of the Root Locus" is used again, but the controller now is incorporated with an integral network instead. The transfer function of such a controller is

$$G_1(s)=K_1\frac{s+\sigma_i}{s+\dfrac{\sigma_i}{\alpha_i}} \tag{33}$$

in which the values of σ_i and α_i are to be selected so that the value K_1 can be increased. The root-locus equation is

$$\text{Ang}\,\frac{1}{s(s+1)}+\text{Ang}\,\frac{s+\sigma_i}{s+\sigma_i/\alpha_i}=\pm 180\text{ degrees} \tag{34}$$

and the magnitude equation is

$$K=\left|s(s+1)(s+\sigma_i/\alpha_i)/(s+\sigma_i)\right| \tag{35}$$

and the velocity constant is

$$K_v=K\alpha_i \tag{36}$$

where $K=K_1K_2$. Thus the velocity constant is increased α_i times. The value α_i is to be selected from the desired increase of the velocity constant subject to the practical limitation. The value σ_i should be so chosen that the root locus will remain nearly the same with the integral compensation. This can be done by means of the phase-angle loci. The phase-angle loci of the system with integral compensation are

$$\text{Ang}\,\frac{1}{s(s+1)}+\text{Ang}\,\frac{s+\sigma_i}{s+\sigma_i/\alpha_i}=\phi \tag{37}$$

By the superposition method, it is split into two expressions

$$\text{Ang}\,\frac{1}{s(s+1)}=\phi_1$$

$$\text{Ang}\,\frac{s+\sigma_i}{s+\sigma_i/\alpha_i}=\phi_i$$

The first expression is that shown in Figure 5. The second one is similar to that shown in Figure 2 but the locations of the pole and zero are different. Their location should be such as to make the root locus almost unchanged. This is done by choosing a small value of σ_i and a large value of α_i and then superimposing the second phase-angle locus on the first one. The result is shown in Figure 17, where the original root locus is little changed except in the neighborhood of the pole and zero of the compensation network. There is about a 5-degree phase shift in the region where the original predominate roots are chosen. It may be concluded that σ_i should be as small as possible and α_i should be as large as possible, but practical limitations exist for both these two values.

Conclusion

By the introduction of phase-angle loci and the superposition method, the elimination of the guesswork of plotting the root locus and the reshaping of the root locus may be achieved. The reshaping also can be achieved by any element of the system, the parameter of which may be varied. The roots chosen from the root locus, together with zeros, will determine the transient response. Here, a synthesis has been formulated so that the transient response can be evaluated readily without its actual calculation. This is possible by developing two approximate equations under a widely held assumption that there exists a pair of predominate complex roots. The two approximate equations also may be used to explain the effect of the location of zeros as well as the other roots on the peak time T_p and the peak overshoot M. Finally, it is shown that for integral compensation the shape of the root locus should be little changed despite the insertion of the integral network.

Appendix I. The Root Locus Method

A brief account of Evans' root-locus method[1] is provided in this appendix. If a feedback control system is given as shown in Figure 14, the system function, given in equation 38, is

$$\frac{C}{R}=\frac{G(s)}{1+G(s)H(s)} \tag{38}$$

The characteristic equation of this system is then

$$G(s)H(s)=-1 \tag{39}$$

where s is the complex variable. This is a complex algebraic equation. It may be separated into two equations by equating the respective magnitudes and phase angles one each side. Thus, the magnitude equation is

$$|G(s)H(s)|=1 \tag{40}$$

and the phase-angle equation is

$$\text{Ang}\,\{G(s)H(s)\}=\pm 180\text{ degrees} \tag{41}$$

Any value of s which fulfills the two equations 40 and 41 is a root of the characteristic equation 39. The plot of the phase-angle equation on the complex s-plane is called the root locus, because every point on this plot is a possible solution (of the characteristic equation). The gain of the loop is employed as a parameter of the root locus and is to be determined from the magnitude equation. Thus the root locus gives a complete picture of all possible roots of the characteristic equation.

The problem of the absolute stability of the system can be solved readily by choosing no roots on the right half of the s-plane. The degree of stability (or relative stability) may be determined by choosing the proper damping ratio of the predominate pair of the complex roots.

However, the plot of the root locus of equation 41 on the s-plane requires some sort of guesswork, except that portion of the root locus which occurs on the real axis and that portion for large values of s. Because of the simplicity of selecting the desirable roots of the characteristic equation, it is a method for synthesis of feedback control system in the time domain.

References

1. Control System Synthesis by Root Locus Method, **Walter R. Evans.** *AIEE Transactions*, volume 69, part I, 1950, pages 66–69.
2. Principles of Servomechanisms, **G. S. Brown, D. P. Campbell.** John Wiley and Sons, Inc., New York, N. Y., 1948.
3. Comparison of Steady-State and Transient Performance of Servomechanisms, **Harold Chestnut, Robert W. Mayer.** *AIEE Transactions*, volume 68, part I, 1949, pages 765–77.
4. Servomechanism Transient Performance from Decibel-Log Frequency Plots, **H. Harris, Jr., M. J. Kirby, E. F. von Arx.** *AIEE Transactions*, volume 70, part II, 1951, pages 1452–59.
5. Quick Methods for Evaluating the Closed-Loop Poles of Feedback Control Systems, **G. A. Biernson.** *Engineering Report Number* 38, Servomechanism Laboratory, Massachusetts Institute of Technology (Cambridge, Mass), March 1952.

No Discussion

Signal Component Control

D. J. GIMPEL
STUDENT MEMBER AIEE

J. F. CALVERT
FELLOW AIEE

THE proposed AIEE symbols and definitions for feedback control systems[1] will be employed throughout this paper except for one omission and two additions. It will be assumed that the command signal is received with suitable magnitude and in appropriate electrical, mechanical, or other physical form so that the "reference input elements" will be omitted, although the term "reference input" will be retained to signify the input signal to the control system. The term "signal component control" (scc) will be introduced to designate the control method, and the term "signal component control device" will be employed to specify the particular associated equipment.

Objective

In Figure 1 it may be assumed that the primary objective of the scc device is to so shape the reference input that, depending upon the designer's choice, the indirectly or directly controlled variable will be at all times nearly identical to the command signal. It is intended to achieve this objective in such a way that:

1. The controlled variable at all times will contain a component which is identical to the command signal.
2. In a very short time (usually this will be made less than one-half the longest natural period of oscillation of the control system), all aperiodic deviations of the controlled variable from the command signal will be reduced to zero.
3. In the same short time each natural mode of the control system will be reduced to zero.

A second objective might be to make the controlled variable follow a derivative of the command function. It is possible to realize this and other objectives as well, but only the primary objective just mentioned will be discussed in this paper.

List of Symbols

Time

t = instantaneous time
T = a period of time in the units of t
T_k = the k^{th} time delay in units of t

Variables

$v(t)$ = command function, a function of time t
$q(t)$ = indirectly controlled variable, a function of t

Parameters

A_0 = a real number
$A_1, A_2, \ldots, A_n$ = real or complex numbers, and, if complex, appearing in conjugate pairs
$B_0, B_1, \ldots, B_n \ldots B_{n+l}$ = real numbers
M_a, M_b, M_c, M_d = real numbers
$\xi_1, \xi_2, \ldots \xi_r \ldots \xi_n$ = real or complex numbers, and, if complex, appearing in conjugate pairs

General Subscripts

$0, 1, 2, \ldots i, k, \ldots, n, \ldots, (n+l)$ = integers

(In the paper n is the order of the differential equation representing the control system and l the degree of the polynomial representing the command signal.)

Theory

Various mathematical functions might be employed to describe the command signal. Polynomial representations will be employed in this discussion. Figure 2(A) shows the command signal $v(t)$ which, as indicated in Figure 2(B), will be defined in the following manner:

In the time interval	$v(t)$ will be characterized by
$t<0$	$v(t)=0$
$0\leqq t\leqq T$	$v(t)=v_1(t)$
$T\leqq t\leqq 2T$	$v(t)=v_1(t)+v_2(t)$
$T\leqq t\leqq 3T$	$v(t)=v_1(t)+v_2(t)+v_3(t)$
	etc.

(1)

In general, the period of time T will be chosen less, and sometimes appreciably less, than one-half the period of the longest natural oscillation of the control system. Let it be expected that the indirectly controlled variable is to match the command signal at the ends of these time periods and that it is not to deviate greatly at any intermediate instant of time. Suppose that at $t=0$, $v_1(t)$ were applied and the controlled variable matched $v_1(t)$ for $t>T$. If (1) this were true regardless of whether the command signal was $v_1(t)$, $v_2(t)$, or $v_3(t)$, and so forth and further if (2) the control system were linear so that superposition principles are applicable, then reference to Figure 2 shows that the desired results may be obtained when these individual signals are applied one after another. Now, let any one of these individual command functions be called $v(t)$, and assume that it can be described for the time $t=0$ to $t=T$ with satisfactory reliability by a polynomial

$$v(t)=M_a+M_bt+M_ct^2+M_dt^3 \qquad 2(A)$$

or perhaps with less accuracy by

Paper 52-211, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952, and approved by the AIEE Committee on Technical Operations for representation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted March 24, 1952; made available for printing August 8, 1952.

D. J. Gimpel and J. F. Calvert are with the Electrical Engineering Department of the Technological Institute, Northwestern University, Evanston, Ill.

The authors wish to acknowledge the helpful criticism of Dr. F. B. Bothwell, J. E. Alexander, and others of the Electrical Engineering staff at Northwestern University.

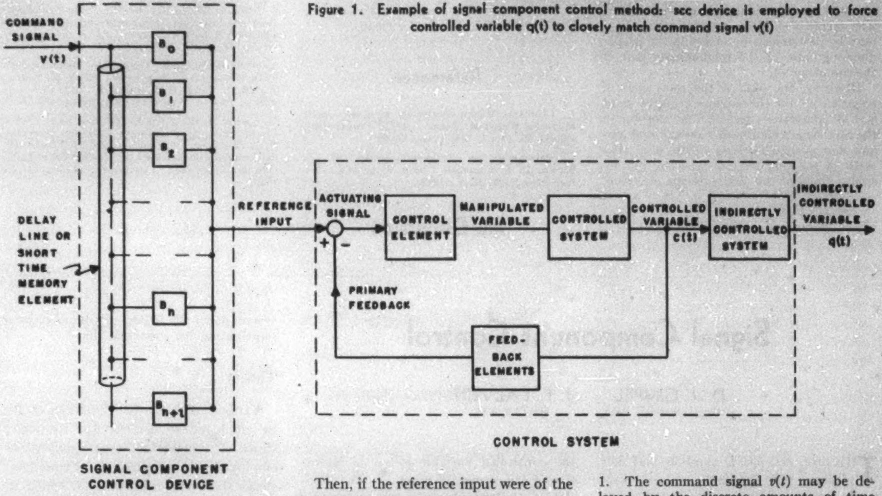

COMMAND SIGNAL
v(t)
B_0
B_1
B_2
B_n
B_{n+1}
DELAY LINE OR SHORT TIME MEMORY ELEMENT
SIGNAL COMPONENT CONTROL DEVICE
REFERENCE INPUT
ACTUATING SIGNAL
+
−
CONTROL ELEMENT
MANIPULATED VARIABLE
CONTROLLED SYSTEM
CONTROLLED VARIABLE
c(t)
INDIRECTLY CONTROLLED SYSTEM
INDIRECTLY CONTROLLED VARIABLE
q(t)
PRIMARY FEEDBACK
FEED-BACK ELEMENTS
CONTROL SYSTEM

Figure 1. Example of signal component control method: scc device is employed to force controlled variable q(t) to closely match command signal v(t)

$$v(t)=M_a+M_bt \tag{2(B)}$$

Although it is not essential to the theory, two simplifying steps will be taken for illustrative purposes. First, equation 2(*B*) will be employed and, second, the control system will be considered to be linear and to have no multiple roots in its associated characteristic equation. Hence, if the reference input applied to the control system itself were a step (or Heaviside unit function) the indirectly controlled variable would be

$$q(t)=A_0+\sum_1^n A_k\epsilon^{\xi kt} \tag{3}$$

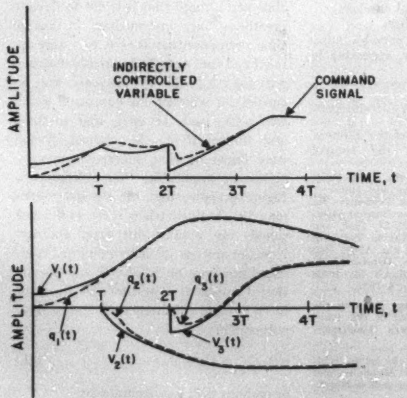


Figure 2. Illustration of superposition of command functions, and of the responses for the representation of the total command signal and the total controlled variable respectively

Then, if the reference input were of the form shown in equation 2(*B*), the superposition theorem may be employed in establishing the response

$$q(t)=M_a\Big[A_0+\sum_1^n A_k\epsilon^{\xi kt}\Big]+\int_0^t M_b[A_0+\sum_1^n A_k\epsilon^{\xi kt}\epsilon^{-\xi k\lambda}]d\lambda \tag{4}$$

$$q(t)=A_0[M_a+M_bt]-M_b\sum_1^n\frac{A_k}{\xi_k}+\sum_1^n A_k\left[M_a+\frac{M_b}{\xi_k}\right]\epsilon^{\xi kt} \tag{5}$$

Assume that the following may be achieved with the scc device:

1. The command signal $v(t)$ may be delayed by the discrete amounts of time $T_1, T_2, \ldots T_{n+1}$, where $0<T_1<T_2<\ldots<T_{n+1}$.
2. At the corresponding instants of time, the command function $v(t)$ may be amplified or attenuated with or without change of sign. That is, $v(t)$ and the delayed forms may be multiplied by the constant positive or negative real numbers $B_0, B_1, B_2 \ldots B_{n+1}$, respectively.
3. The results of these two operations are summed to yield the output of the scc device, which is the reference input applied to the control system.

Under these conditions the response of the control system becomes, for $t>T_{n+1}$

$$\begin{aligned}q(t)={}&B_0A_0[M_a+M_bt]-B_0M_b\sum_1^n\frac{A_k}{\xi_k}+\\&B_0\sum_1^n A_k\left(M_a+\frac{M_b}{\xi_k}\right)\epsilon^{\xi kt}+\\&B_1A_0[M_a+M_bt]-B_1M_b\sum_1^n\frac{A_k}{\xi_k}+\\&B_1\sum_1^n A_k\left(M_a+\frac{M_b}{\xi_k}\right)\epsilon^{\xi k(t-T_1)}-\\&B_1M_bA_0T_1+\\&\vdots\\&B_{n+1}A_0[M_a+M_bt]-B_{n+1}M_b\sum_1^n\frac{A_k}{\xi_k}+\\&B_{n+1}\sum_1^n A_k\left(M_a+\frac{M_b}{\xi_k}\right)\times\\&\epsilon^{\xi k(t-T_{n+1})}-B_{n+1}M_bA_0T_{n+1}\end{aligned} \tag{6}$$

Any of the B and the T values may be varied within broad limits. Hence, it may be elected to place restrictions on equation 6.

Accordingly, where it is assumed that $A_0 \neq 0$, let

$$A_0[M_a+M_bt][B_0+B_1+\ldots+B_{n+l}] = M_a+M_bt \qquad 7(A)$$

$$-M_b\left[\left(\sum_1^n \frac{A_k}{\xi_k}\right)B_0+\left(A_0T_1+\sum_1^n \frac{A_k}{\xi_k}\right)\times B_1+\ldots+\left(A_0T_{n+l}+\sum_1^n \frac{A_\kappa}{\xi_k}\right)B_{n+l}\right]=0 \qquad 7(B)$$

$$\left.\begin{array}{l} A_1\left[M_a+\dfrac{M_b}{\xi_1}\right]\epsilon^{\xi_1 t_1}[B_0+\epsilon^{-\xi_1 t_1}B_1+\ldots+ \epsilon^{-\xi_1 T_{n+l}}B_{n+l}]=0 \\ \cdot \\ \cdot \\ \cdot \\ A_n\left[M_a+\dfrac{M_b}{\xi_n}\right]\epsilon^{\xi_n t}[B_0+\epsilon^{-\xi_n T_1}B_1+ \ldots+\epsilon^{-\xi_n T_{n+1}}B_{n+l}]=0 \end{array}\right\} 7(C)$$

From equations 7

$$B_0+B_1+B_2+\ldots+B_{n+l}=\frac{1}{A_0} \qquad 8(A)$$

$$T_1B_1+T_2B_2+\ldots+T_{n+l}B_{n+l}=\frac{1}{A_0^2}\sum_1^n \frac{A_R}{\xi_R} \qquad 8(B)$$

$$\left.\begin{array}{l} B_0+\epsilon^{-\xi_1T_1}B_1+\epsilon^{-\xi_1T_2}B_2+\ldots+ \epsilon^{-\xi_1T_{n+l}}B_{n+l}=0 \\ B_0+\epsilon^{-\xi_nT_1}B_1+\epsilon^{-\xi_nT_2}B_2+\ldots+ \epsilon^{-\xi_nT_{n+l}}B_{n+l}=0 \end{array}\right\} 8(C)$$

The B and T values, as determined from equations 8, define a set of signal components which, when employed in the design of the scc device, permit the primary objective to be achieved. It will be noted that while equations 8 are influenced by the form chosen for $v(t)$, the magnitudes of the coefficients in $v(t)$ and the independent variable t no longer appear. It will be elected to make l the degree of the polynomial representing the command signal. In equations 8 this will yield $(n+2)$ equations containing $(n+2)$ of the B terms and $(n+1)$ of the T terms. Equations 8 are linear with respect to the former and nonlinear with respect to the latter terms. In consequence, if $(n+l+1)$ is large, it is far easier to choose values for the T terms and then compute the B values than it is to carry out the reverse of this process. The T values, of course, are positive real numbers. It may be noted that the ξ terms and the A terms appear either as real numbers or else in complex conjugate pairs. In every instance the B terms will be either positive or negative real numbers.

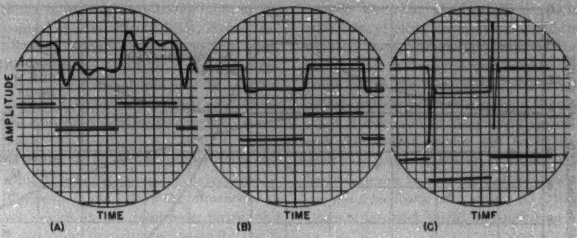


Figure 3. Response of a 2-mode oscillatory control system to a step command signal

All lower curves are commands. Upper curves of (A) and (B) show controlled variable without and with signal component control respectively. Upper curve of (C) is reference input for case (B) where scc device is employed

In the development of equations 8, the general form of equation 2(B) described both the initial command signal and the changes in the shape of the command signal across the succeeding time periods. If additional terms are employed in the polynomial, as shown, for example, by equation 2(A), then an equation similar to 6 is obtained. Within this new equation the choice of the B and T values may be varied and, as before, restrictions may be established in order to achieve the primary objective of the scc method. Thus, similar to equations 8, the following will be obtained:

1. There will be a single equation as in 8(A) which specifies that a component of the control variable at all times matches the command signal;
2. There will be instead of the single equation 8(B) a set of equations which require that all aperiodic output errors be eliminated in a short period of time;
3. There will be a set of equations which, like 8(C), require that the natural modes appearing in the controlled variable cease in the same short period of time.

Again, this time may be chosen less than the longest natural period of oscillation for the control system. It will be found that the magnitudes of the coefficients M_a, M_b, M_c, M_d, ... of equation 2(A) do not appear in this last system of equations which is similar in requirements to those of equations 8. Thus the scc method may employ theoretical criteria which are independent of the coefficients of the polynomial representing the command signal. To meet the primary objective, the theoretical criteria for the design of the scc device depend only on

1. The degree of the polynomial employed in the representation of the command signal, and
2. The actual parameters employed in the linear representation of the control system.

Thus far, it has been assumed that the signal component control device was placed between the location where the command signal was received and the location where the reference input was delivered to the control system. Equations 6, 7, and 8 show that, when the control system may be treated as linear, the foregoing does not represent the only possible arrangement of the equipment. For example, the same equations and hence the same theoretical criteria would have been developed if Figure 1 were redrawn such that the scc device came after the control system instead of ahead of it. In this case the output of the scc device would be the controlled variable $q(t)$. If instead of using the superposition theorem both the performance of the scc device and of the linear control system had been represented by linear operators, it would

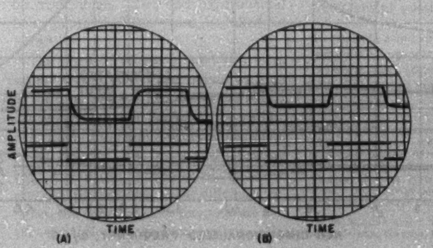


Figure 4. Response of a single mode control system to a step command signal

Lower curves are commands. Upper curves show controlled variable without and with signal component control respectively

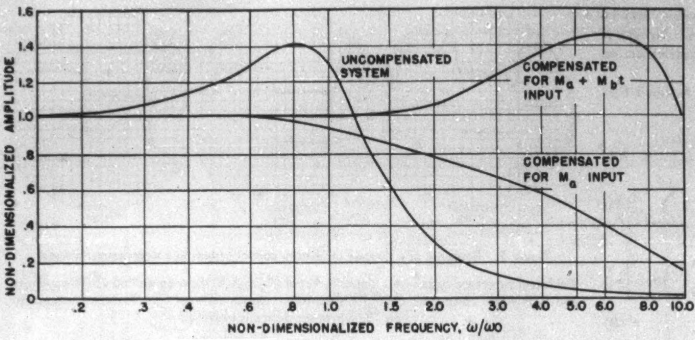


Figure 5. Amplitude response corresponding to control system of Figure 3

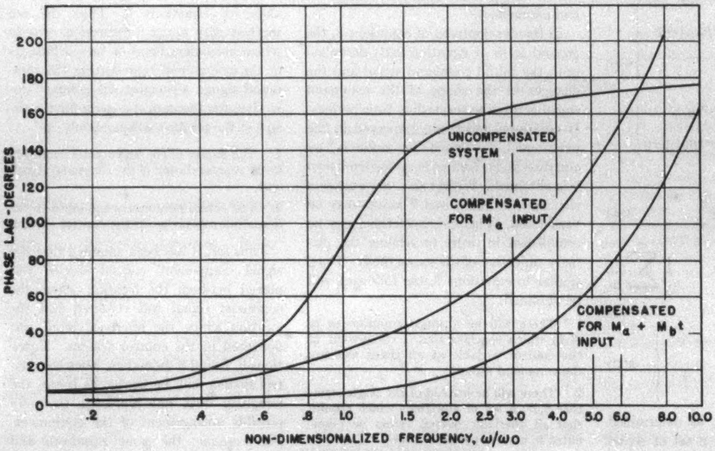


Figure 6. Phase lag response corresponding to control system of Figure 3

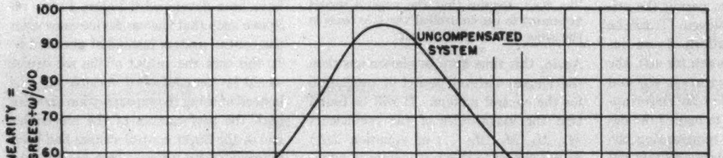


Figure 7. Phase delay linearity corresponding to control system of Figure 3

have been observed that the latter were commutative.

Experimental Results

Experimental data will be presented to illustrate some of the results which may be expected from the use of the scc method. Two linear control systems are selected. One is a first-order and the other a second-order oscillatory system. The second-order system had negative feedback included within it. Two forms of the scc device are used. One is designed for only a step input M_a and the other for an input which includes a step plus a straight line term $[M_a + M_b t]$. All tests employed a delay line and were carried out at audio frequencies. The controlled and indirectly controlled variables are identical. Figure 3 applies to the control system having two natural modes. The controlled variable follows the step signal, when the scc device is employed, in considerably less than one-half the natural period of the control system. However, the output from the scc device which is, in fact, the reference input to the control system, Figure 3(C), shows momentary overshoots that exceed the magnitude of the step-type command function. These overshoots may be avoided at some sacrifice in the speed with which the controlled variable reaches the desired value.

Figure 4 illustrates similar data for a control system having one natural mode. Figures 5, 6, and 7 show the influence of an scc device on the amplitude, phase, and phase-delay linearity responses for the 2-mode control system of Figure 3. As used here, the term phase-delay linearity means phase angle divided by the nondimensionalized angular rate for an oscillatory system of two natural modes.

Summary

The primary objective of the signal component control method is to compel either the directly or indirectly controlled variable (as may be desired) to follow more closely the command signal. The control system and the scc device may contain electrical, mechanical, or parts which are of other physical form. In its simplest form, the method involves the use of an scc device employed between the point where the command signal is received and the point where the reference input is delivered to the control system, and the control system will be of the open-loop type. However, it usually appears desirable to have the control system contain a feedback path within it. Other types of feedback might be considered:

(1). an integral type, placed around the entire system to relate $q(t)$ and $v(t)$ for the purpose of preventing drift errors, and

(2). feedback from a variable parameter (or the load treated as a variable parameter) to so alter the B values in the scc device that they could be adjusting continually to meet the requirements of the control system.

References

1. PROPOSED SYMBOLS AND TERMS FOR FEEDBACK CONTROL SYSTEMS, **AIEE Committee Report.** *Electrical Engineering*, volume 70, number 10, October 1951, pages 905–09.

2. A PHASE-PLANE APPROACH TO THE COMPENSATION OF SATURATING SERVOMECHANISMS, **Arthur M. Hopkin.** *AIEE Transactions*, volume 70, part I, 1951, pages 631–39.

No Discussion

Residential Heat Pump Experiments in Philadelphia—Earth as a Heat Source

A. H. KIDDER
FELLOW AIEE

J. H. NEHER
MEMBER AIEE

Synopsis: This paper summarizes ground coil performance during two heating seasons at Whitemarsh and Lansdowne. There are also derived certain approximate relationships which should be useful for designing ground coils for other situations.

EARTH appears to be superior to air or water as a source for heat pumps operating to provide year-round air comfort in the home. Its ambient temperature is higher and less variable than outdoor air, and its use does not involve problems of accessibility of recirculation such as would accompany the use of surface or underground water. However, little factual information has been available on the heat transfer characteristics of the earth, particularly as a heat source. Also, some investigators were apprehensive that prolonged heat withdrawal or absorption would have adverse effects upon the long-term temperature levels in the earth. From considerations of weather conditions in Philadelphia and the existing state of development of the art, the Philadelphia Electric Company undertook a research project in 1947 on the heat transfer features of the earth to obtain as much as possible of factual information on the use of the earth as a heat or heat sink for heat pumps. Two such experimental heat pump systems were installed in 1948.[1]

The purpose of the ground coil is to supply its associated heat pump with all of the heat which, when supplemented by the energy required for pump operation, will satisfy the requirements of the space to be heated. The total space heating requirement can be estimated, being essentially proportional to the difference between the desired indoor temperature and the variable outdoor temperature. Experience has confirmed the assumption that the required ground coil output is very nearly two-thirds of the total space heating requirement. The minimum permissible ground coil temperature is determined by the minimum suction temperature at which the heat pump will deliver its required output on the coldest day, after accounting for temperature drop in the heat exchanger. It should be noted that the minimum practical suction temperature for the heat pump may be substantially lower than the corresponding ground coil temperature, depending upon whether an intervening heat exchanger is used and its size.

The heat to be extracted by the ground coil has come from the sun at one time or another and is available in one or more of the following possible forms:

1. Simple conduction from the surrounding earth, proportional to the difference between the ambient earth temperature and the average coil temperature.

Paper 52-341, recommended by the AIEE Domestic and Commercial Applications Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted January 15, 1952; made available for printing September 3, 1952.

A. H. KIDDER and J. H. NEHER are both with the Philadelphia Electric Company, Philadelphia, Pa.

Table I. Representative Values of Empirical Constant E for Earth as a Source Derived from Data Summarized in Tables II and III

	Average Q/L	Average (T_a-T_c)	Coil Temperature		Representative Value of E
			Average	Minimum	
Maximum 3 months ended					
Whitemarsh 3/23/49	14.0	12.1	35.2	28.0	1.39
Whitemarsh 3/1/50	27.6	17.9	31.9	26.0	1.85
Lansdowne 3/23/49	19.5	14.8	33.8	28±	1.58
Lansdowne 2/28/50	26.2	19.6	30.0	27.0	1.60
Representative value					1.6
Maximum 4 weeks, ended					
Whitemarsh 3/23/49	23.9	15.1	30.8	28.0	1.90
Whitemarsh 3/1/50	33.5	18.1	29.9	27.0	2.22
Lansdowne 3/23/49	21.2	15.5	32.1	28±	1.64
Lansdowne 2/28/50	30.2	18.6	29.3	27.0	1.95
Representative value					1.9
Maximum week, ended					
Whitemarsh 3/16/49	27.0	16.3	29.6	28.0	1.99
Whitemarsh 3/1/50	36.7	18.7	28.0	27.0	2.35
Lansdowne 3/9/49	28.0	16.0	31.5	28±	2.10
Lansdowne 2/28/50	35.0	20.0	28.0	27.0	2.10
Representative value					2.1
Maximum day					
Whitemarsh 2/20/50	42.7	19.0	28.0	28.0	2.70
Lansdowne 2/20/50	40.4	21.0	27.0	27.0	2.30
Representative value					2.5

2. Specific heat of the earth, jointly proportional to the specific heat, the density, and the temperature change induced by the ground coil.

3. Latent heat of freezing the earth near the ground coil.

4. Some supplementary heat transfer effects which also may accompany the induced temperature changes in any small quantities of water and water vapor which may migrate toward the coil, and then drain away gravitationally, if permitted.

Where there is no drainage, moisture can migrate only for the short time required to produce steady-state saturation of the earth near the cool surface. Where there is drainage, however, moisture migration in either the vapor or the liquid phase should be a more or less continuous process and this effect should appear in the form of an otherwise unexplained high rate of heat extraction from porous or loosely packed earth in a well-drained location.

On the other hand, the specific heat of the earth and the latent heat of freezing produce effects comparable to a heat reservoir which is alternately charged and discharged from time to time during the course of normal variations above and below the "base load" or average rate of heat extraction.

The over-all effect of all the factors is that the total heat required is extracted by what amounts to the equivalent of simple conduction at a relatively stable average rate and average coil temperature, while the transient variations above and below the average rate are offset by the heat reservoir effects in the immediate vicinity of the pipe. This is especially characteristic of operations with coil temperatures near 32 degrees Fahrenheit, because alternate freezing and thawing greatly amplify the heat storage effects.

Calculation of the temperature differentials required to maintain a given pattern of heat extraction rates by use of transient heat flow theory is laborious even for operations which do not involve freezing. Such calculations have been carried out

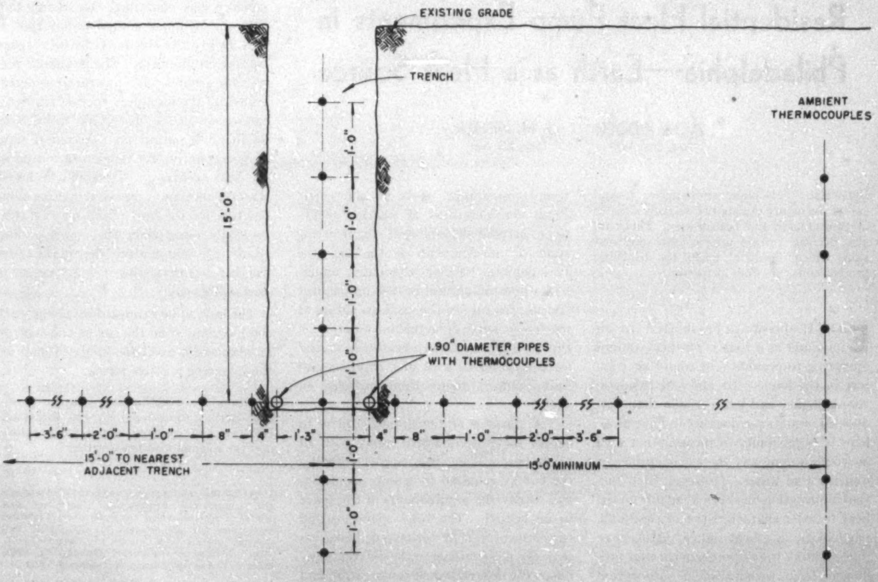


Figure 1. Cross section of ground coil showing location of pipes and thermocouples

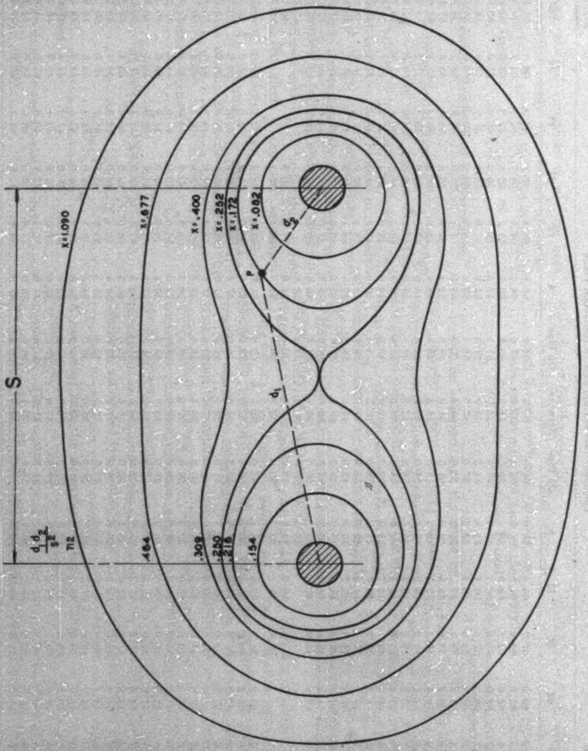


Figure 2. Isotherms for a 2-pipe ground coil. Each isotherm is the locus of points for which d_1d_2/S^2 is a constant and the enclosed area per pipe is XS^2 where the value of X is given for each isotherm. The equivalent radius is equal to d_1d_2/S

Table II. Whitemarsh Coil Operation

Weeks	Period Ended	Average Earth Ambient, Degrees Fahrenheit	Average Coil Temperature, Degrees Fahrenheit	Coil Output in Thousand Btu's	Coil Length in Feet	Hours in Test Period	Coil Output (Btu)/(Hour)(Foot)
1	10/13/48	64.6	64.1	769	1,400	168	3.3
2	10/20/48	63.3	60.4	1,612	1,400	168	6.9
3	10/27/48	61.6	56.0	1,981	1,400	168	8.2
4	11/3/48	58.9	54.9	1,426	1,400	168	6.1
5	11/10/48	58.9	57.1	976	1,400	168	4.0
6	11/17/48	58.6	53.4	1,654	1,400	168	7.0
7	11/24/48	57.3	50.3	2,116	1,400	168	9.0
8	12/1/48	55.3	45.3	2.955	1,400	168	12.6
9	12/8/48	54.1	43.6	2,755	1,400	168	11.7
10	12/15/48	51.7	39.7	3,424	1,400	168	14.6
11	12/22/48	49.7	35.9	3,900	1,550	168	15.5
12	12/29/48	48.9	38.1	4,220	2,220	168	11.4
13	1/5/49	48.6	39.1	3,465	2,220	168	9.3
14	1/12/49	48.3	39.1	3,036	2,220	168	8.2
15	1/19/49	48.3	39.1	3,077	2,220	168	8.3
16	1/26/49	48.4	40.1	2,827	2,220	168	7.6
17	2/2/49	47.1	37.4	3,882	2 220	168	10.5
18	2/9/49	46.4	35.7	3,473	2,030	168	10.2
19	2/16/49	46.1	32.0	2,449	1,280	168	11.4
20	2/23/49	46.3	31.3	2,951	1,280	168	13.7
21	3/2/49	46.3	32.0	3,132	770	168	24.3
22	3/9/49	46.1	31.3	2,910	770	168	22.6
23	3/16/49	45.9	29.6	3.482	770	168	27.0
24	3/23/49	45.4	30.4	2,810	770	168	21.8
25	3/30/49	45.6	34.1	294	770	168	2.0
26	4/6/49	47.9	32.6	2,143	825	168	15.5
27	4/13/49	49.0	35.4	1,708	1,400	168	7.3
28	4/20/49	50.1	36.3	1,660	1,400	168	7.1
29	4/27/49	51.3	41.3	1,025	1,400	168	4.4
1	9/28/49	64.7	66.0	142	770	72	2.5
2	10/5/49	63.4	64.4	443	770	168	3.4
3	10/12/49	62.7	70.3	−793	770	168	−6.1
4	10/19/49	62.0	63.1	784	770	168	6.1
5	10/26/49	61.6	58.0	1,060	770	168	8.2
6	11/2/49	60.3	51.9	1,810	770	168	14.0
7	11/9/49	58.4	44.4	1,983	770	168	15.4
8	11/16/49	56.7	52.6	1,151	770	168	8.9
9	11/23/49	56.0	37.9	3,488	770	168	27.9
10	11/30/49	53.7	32.1	3,570	770	168	27.7
11	12/7/49	51.6	30.6	3,870	770	168	30.0
12	12/4/49	49.4	29.9	3,560	770	168	27.6
13	12/21/49	48.0	29.6	3,229	770	168	25.0
14	12/28/49	47.7	30.7	2,558	770	168	19.8
15	1/4/50	50.1	33.0	2,845	770	168	22.1
16	1/11/50	50.3	33.9	2.758	770	168	21.4
17	1/18/50	50.4	33.1	3.119	770	168	24.2
18	1/25/50	50.0	32.7	3,630	770	168	28.2
19	2/1/50	49.9	33.1	3,285	770	168	27.5
20	2/8/50	49.4	30.9	4,269	770	168	33.1
21	2/15/50	48.4	31.1	3,735	770	168	29.0
22	2/21/50	47.4	29.7	4,606	770	168	35.0
23	3/1/50	46.7	28.0	4,737	770	168	36.7
24	3/7/50	46.0	28.8	3.471	770	144	31.3
Maximum Day	2/20/50	47	28	790	770	24	42.7

far enough for the unfrozen condition, however, to show that the temperature drop from ambient earth to ground coil can be calculated directly, in terms of a given thermal conductivity, thermal diffusivity, and changing rates of heat extraction. When the temperature of the ground coil is reduced below freezing, however, certain additional effects so complicate the problem that an adequate evaluation of the coefficients involved becomes virtually impossible. These effects are discussed later.

It has become apparent from analysis of temperatures above freezing that the average rate of heat extraction soon stabilizes at a value which is approximately proportional to the temperature difference between earth ambient and ground coil. This suggests that no important error should be introduced by treating the heat extraction rate tentatively, as if the temperature change corresponding to a given average rate of heat extraction reaches its steady-state value within approximately a week. Dealing tentatively in terms of steady-state conditions in the earth permits materially simplifying the analysis of the coil temperature changes which accompany the comparatively slow-mov-

Table III. Lansdowne Coil Operation

Weeks	Week Ended	Average Earth Ambient, Deg. F.	Average Coil Temperature, Deg. F.	Coil Output in Thousand Btu's	Coil Length in Feet	Hours in Test Period	Coil Output (Btu)/(Hour)(Foot)
1	10/12/48	62	62	450	860	168	3.1
2	10/19/48	60	60	465	860	168	3.2
3	10/26/48	60	58	947	860	168	6.6
4	11/2/48	59	56	644	860	168	4.5
5	11/9/48	58	56	508	860	168	3.5
6	11/16/48	58	53	1,070	860	168	7.4
7	11/23/48	56.5	51	873	860	168	6.0
8	11/30/48	56.5	49	1,300	860	168	9.0
9	12/7/48	57.5	48	1,320	860	168	9.1
10	12/14/48	56	46	1,490	860	168	10.3
11	12/21/48	55	41	2,420	860	168	16.7
12	12/28/48	53	36	3,000	860	168	20.8
13	1/4/49	51	36	2,450	860	168	23.9
14	1/11/49	49	38	1,500	860	168	10.4
15	1/18/49	48.5	36	2,150	860	168	14.9
16	1/25/49	48.5	36	1,880	860	168	13.0
17	2/1/49	48.5	35	2,570	860	168	17.8
18	2/8/49	48.5	35	2,500	860	168	17.3
20	2/23/49	47	32	3,400	620(average)	360	15.2
22	3/9/49	47.5	31.5	4,230	450	336	28.0
24	3/23/49	47.5	30	3,680	450	336	24.4
26	4/6/49	47.5	34	1,740	450	336	11.5
28	4/20/49	50	40	2,640	860	336	9.1
4	10/18/49	64.5	64	226	650(average)	672	0.5
8	11/15/49	63	55	5,040	450	672	16.6
10	11/29/49	59.5	46	4,680	450	336	31.0
11	12/6/49	56.5	34	2,195	450	168	29.0
12	12/13/49	54.5	31	2,120	450	168	28.3
13	12/20/49	53	31	2,370	450	168	31.4
14	12/27/49	51	31	1,625	450	168	21.6
15	1/3/50	49	30	1,805	450	168	23.9
16	1/10/50	49.5	30	1,210	450	168	16.0
17	1/17/50	50	30	1,740	450	168	23.1
18	1/24/50	49	29	2,130	450	168	28.2
19	1/31/50	47.5	31	1,395	450	168	18.5
20	2/7/50	47.5	29	2,255	450	168	29.9
21	2/14/50	48.5	30	2,015	450	168	26.7
22	2/21/50	47.5	30	2,195	450	168	29.1
23	2/28/50	48	28	2,640	450	168	35.0
Maximum Day	2/20/50	48	27	436	450	24	40.4

Table IV. Thermocouple Readings at Whitemarsh in Degrees Fahrenheit

	Left				Location—Inches from Pipe in Horizontal Plane									Right
Date	90	48	24	12	4	On Pipe	7½ (Between)	On Pipe	4	12	24	48	90	180
10/6/48	68.9	68.0	67.7	66.6	65.0	64.0	64.4	64.6	65.8	66.8	67.8	67.0	66.6	65.5
10/13/48	66.5	65.8	65.2	64.5	63.6	62.0	62.9	62.6	64.4	64.6	64.7	64.4	64.4	63.5
10/20/48	65.4	64.6	63.5	62.5	60.1	59.0	58.3	58.1	60.7	61.7	63.1	62.8	62.6	62.0
10/27/48	63.3	62.0	60.8	60.1	57.3	57.5	55.7	57.2	57.5	58.6	59.9	60.4	61.0	60.0
11/3/48	59.7	58.3	57.6	56.1	54.8	54.5	55.0	54.3	54.8	55.8	56.8	57.7	57.9	58.0
11/10/48	60.7	60.4	59.5	59.1	58.2	57.0	56.9	58.1	58.1	58.8	59.9	59.7	60.0	59.0
11/17/48	60.5	59.8	59.1	57.7	55.3	53.0	54.3	50.0	55.0	57.1	57.9	57.9	58.9	59.0
11/24/48	57.8	56.8	56.0	55.2	52.0	51.0	51.6	49.8	52.4	54.8	56.8	55.6	56.0	57.0
12/1/48	56.8	56.3	53.3	52.3	48.3	45.5	48.3	46.6	48.8	52.3	54.6	55.4	55.8	56.0
12/8/48	50.8	51.3	50.0	49.2	46.4	44.0	45.6	42.8	45.6	47.3	50.0	50.4	51.3	53.0
12/15/48	50.7	50.3	48.0	46.3	43.5	38.0	43.5	39.1	43.5	46.3	48.5	48.5	50.6	51.0
12/22/48	48.8	47.4	45.6	42.4	39.9	35.5	38.0	35.4	40.9	42.4	44.4	45.6	47.4	49.0
12/29/48	48.3	47.4	45.1	44.7	41.9	39.5	42.9	40.3	42.9	43.9	46.4	47.4	49.1	49.5
1/5/49	46.4	45.4	43.7	43.7	41.9	38.0	40.5	38.4	40.9	43.3	44.5	44.9	47.3	48.0
1/12/49	46.2	46.2	45.4	44.4	42.6	38.0	41.8	38.9	42.6	43.4	43.9	44.4	46.2	48.0
1/19/49	47.7	47.3	44.9	43.7	43.3	42.0	41.3	38.8	43.3	44.1	44.9	45.8	47.7	49.0
1/26/49	44.4	44.4	44.4	43.6	43.6	39.0	40.4	37.6	43.6	43.6	43.6	44.0	45.2	47.0
2/2/49	45.8	44.8	43.6	42.3	40.3	35.5	38.0	35.3	49.3	43.2	45.4	44.0	46.3	47.0
2/9/49	43.8	43.8	43.0	40.6	39.1	36.5	39.1	37.4	38.6	42.1	43.0	43.4	45.1	46.0
3/9/49					33.0	30.0	32.5	29.4	33.8					45.0
3/23/49			40.8	38.0	34.8	31.0	33.2	30.4	34.0	37.2	40.8			46.0
4/6/49			38.6	34.1	30.9	30.9	30.9	30.9	31.3	35.1	39.1			48.4
10/5/49	65.6	65.2	63.5	62.5	64.4	65.0	62.5	65.6	64.0	63.5	64.4	62.5	62.0	62.0
10/19/49	66.0	65.2	63.5	63.5	63.5	58.5	62.5	59.0	63.5	64.0	65.2	64.0	63.5	62.5
11/2/49	63.0	62.0	61.6	59.0	55.0	46.5	55.5	46.0	57.6	60.0	62.0	62.0	62.0	60.4
11/16/49	57.6	57.6	55.0	53.2	50.5	51.0	50.4	51.0	51.0	55.0	55.0	55.0	55.0	56.0
11/30/49	50.0	49.2	45.6	40.8	36.0	31.0	37.2	30.8	37.2	43.0	46.0	48.8	50.0	52.8
12/7/49	48.8	47.0	44.0	39.5	34.5	29.0	35.5	28.5	34.5	40.0	44.8	46.5	48.8	50.5
12/14/49	45.2	43.5	40.0	36.0	32.5	29.0	32.0	28.5	32.5	36.8	40.8	44.0	45.2	47.5
12/21/49	43.0	41.6	38.0	34.0	31.6	28.0	30.8	28.0	32.0	35.5	40.0	41.6	44.4	46.5
12/28/49	45.6	44.8	42.0	38.0	36.0	33.0	35.5	33.5	36.4	40.0	44.0	44.8	48.0	49.2
1/4/50	46.0	45.6	44.0	39.5	37.2	34.0	37.2	34.5	37.2	44.8	44.4	46.0	48.4	50.0
1/11/50	45.2	44.8	41.2	37.6	35.5	32.0	35.5	31.5	35.5	39.5	42.0	44.4	47.0	50.0
1/18/50	45.2	44.8	41.6	37.6	35.5	32.0	35.5	32.0	35.5	40.0	43.0	45.2	48.4	49.2
1/25/50	45.6	44.4	42.0	39.5	36.8	33.0	37.2	33.0	36.8	40.4	43.0	45.6	48.4	49.2
2/1/50	44.8	44.4	40.8	39.0	35.5	33.0	35.5	30.8	35.5	39.5	43.0	45.2	46.5	50.0
2/8/50	44.0	44.4	40.8	37.2	35.5	29.5	33.0	30.0	34.0	38.5	42.0	44.4	45.6	48.8
2/15/50	42.0	41.6	40.0	36.4	33.0	31.0	32.5	31.2	34.0	37.2	40.0	43.5	44.8	48.0
2/21/50	40.8	40.4	38.5	35.5	32.5	27.0	30.8	27.2	32.0	36.8	40.0	41.2	44.4	47.6
3/1/50	41.6	41.2	39.5	36.4	32.5	27.5	33.0	27.6	35.0	37.6	41.2	44.4	45.6	46.5
3/8/50	40.8	40.0	36.8	34.0	32.5	31.5	31.2	31.6	31.6	35.5	37.6	40.0	43.0	44.8
3/22/50	40.8	40.0	37.2	35.5	36.0	35.5	36.4	36.0	36.0	36.8	40.0	40.0	44.0	44.4
4/5/50	43.5	41.6	41.2	39.0	36.8		36.4	36.8	37.2	38.5	41.6	44.4	45.6	47.5

Table V. Thermocouple Readings at Whitemarsh in Degrees Fahrenheit

	Location—Feet Below Ground Level											
	Center Line of Trench						Ambient Earth					
Date	2	3	4	5	6	7	2	3	4	5	6	7
10/6/48	65.3	66.4	66.2	64.4	64.2	64.2	63.7	66.6	66.4	65.5	65.1	64.0
10/13/48	62.7	63.2	63.4	62.9	62.6	62.0	62.3	63.4	64.2	63.5	63.2	63.2
10/20/48	57.8	59.6	59.6	58.3	60.1	60.5	57.8	60.8	61.9	62.0	62.5	62.8
10/27/48	55.0	55.7	55.9	55.7	57.5	58.1	56.3	57.9	59.9	60.0	61.1	61.0
11/3/48	54.0	54.3	55.0	55.0	56.8	57.6	54.3	55.9	56.7	58.0	58.4	58.4
11/10/48	57.7	57.3	57.1	56.9	57.5	58.1	58.8	59.5	59.9	59.0	60.0	60.6
11/17/48	52.6	54.4	54.3	54.3	55.3	56.7	54.3	57.1	58.6	59.0	59.8	60.5
11/24/48	50.0	52.0	51.6	51.6	52.8	56.0	53.3	56.0	55.6	57.0	57.3	59.3
12/1/48	47.4	50.2	50.2	48.3	52.8	55.4	49.3	52.8	55.4	56.0	57.3	58.6
12/8/48	44.8	44.8	43.8	45.6	49.2	50.8	47.3	49.6	51.8	53.0	52.1	54.4
12/15/48	42.7	43.5	43.5	43.5	47.1	48.5	45.9	47.0	49.9	51.0	52.0	54.3
12/22/48	38.0	38.4	38.0	38.0	43.4	46.4	42.0	45.2	47.9	49.0	50.4	52.4
12/29/48	40.3	40.7	42.9	42.9	47.4	49.9	41.5	44.7	47.9	49.5	51.9	53.5
1/5/49	39.7	40.1	40.5	40.5	45.4	48.5	41.9	44.1	47.3	48.0	48.9	51.7
1/12/49	41.8	40.9	41.8	41.8	44.4	47.0	46.2	45.4	47.0	48.0	48.4	51.4
1/19/49	40.9	40.9	41.3	41.3	45.3	47.7	44.5	46.8	47.7	49.0	49.3	52.1
1/26/49	40.4	40.4	40.0	40.4	44.0	44.8	43.6	44.8	45.6	47.0	48.8	48.8
2/2/49	39.3	39.3	39.3	38.0	43.2	46.3	42.3	44.0	46.8	47.0	48.0	49.3
2/9/49	39.1	39.1	39.1	39.1	43.0	44.2	40.6	43.0	43.8	46.0	47.4	49.6
3/9/49				32.5						45.0		
3/23/49			33.6	33.2	35.8					46.0		
4/6/49			34.6	30.9	34.1					48.4		
10/5/49	62.5	62.5	64.0	62.5	64.0	62.5	61.6	62.0	62.5	62.0	62.0	60.0
10/19/49	63.5	65.2	64.0	62.5	63.5	63.5	62.0	63.5	63.0	62.5	62.0	60.8
11/2/49	58.5	59.5	57.6	55.5	59.0	59.5	58.0	60.4	61.2	60.4	61.2	61.6
11/16/49	52.8	52.8	52.8	50.4	54.5	55.0	54.0	55.0	55.5	56.0	57.2	57.6
11/30/49	40.0	40.8	39.5	37.2	45.2	49.6	41.6	45.2	48.8	52.8	53.2	54.0
12/7/49	37.2	38.5	36.8	35.5	41.6	48.8	40.0	44.0	46.5	50.5	51.0	53.2
12/14/49	36.8	35.5	33.0	32.0	39.5	45.6	41.2	42.0	44.5	47.5	48.8	50.5
12/21/49	34.0	35.0	32.0	30.8	37.6	44.0	40.0	40.8	43.5	46.5	48.4	50.0
12/28/49	42.0	39.5	37.2	35.5	41.6	46.5	47.5	46.0	48.8	49.2	51.0	53.2
1/4/50	41.6	40.0	37.6	37.2	42.0	46.5	46.5	46.0	48.8	50.0	52.0	53.2
1/11/50	41.2	40.0	36.8	35.5	40.4	45.2	45.2	46.0	47.5	49.6	50.5	51.0
1/18/50	40.8	40.0	36.8	35.5	40.4	45.2	45.2	48.4	48.8	49.2	50.5	52.8
1/24/50	40.4	40.0	37.6	37.2	41.2	46.0	46.0	46.5	48.4	49.2	51.0	53.2
2/1/50	41.6	40.0	37.2	35.5	40.0	44.8	46.5	48.8	48.8	50.0	50.0	52.0
2/8/50	37.6	37.6	36.0	33.0	38.0	43.5	42.5	45.2	46.0	48.8	49.2	50.5
2/15/50	37.2	37.2	35.0	32.5	36.4	42.0	42.0	44.4	45.6	48.0	48.8	50.0
2/21/50	36.8	36.4	35.0	30.8	36.8	41.2	40.4	41.6	44.8	47.6	48.4	49.2
3/1/50	36.8	38.0	37.2	33.0	37.6	44.8	40.4	43.0	44.8	46.5	48.4	50.0
3/8/50	36.0	36.0	34.0	31.2	34.6	40.8	37.6	40.4	41.6	44.8	45.1	48.0
3/22/50	36.8	36.8	36.8	36.4	38.5	43.5	41.6	42.0	43.5	44.4	46.0	48.4
4/5/50	45.6	40.0	37.2	36.4	40.0	43.5	52.0	48.0	48.8	47.5	48.0	48.8

ing seasonal variations in average heat extraction rates. The validity of the assumption then may be checked against test results.

Empirical Analysis of Performance

With incidental transient heat storage effects of the earth environment thus tentatively disregarded, analysis of the heat extraction rates reduces to consideration of but two independent factors which are respectively designated as a "geometric factor" and an "empirical earth conductivity." Under these conditions the effects of pipe size, separation between adjacent pipes, and burial depth can be accounted for approximately in terms of the Kennelly formula for the steady state,[2] for which the steady-state geometric factor is

$$G=0.366\left(\log_{10}\frac{2H}{R}+\log_{10}\frac{\sqrt{4H^2+S_1^2}}{S_1}+\dots+\log_{10}\frac{\sqrt{4H^2+S_n^2}}{S_n}\right) \qquad (1)$$

where

H = depth of the ground coil below earth surface
R = radius of the reference pipe
$S_1 \dots S_n$ = the respective radial distance from the reference pipe (usually one of the middle pipes) to each of the n nearby parallel pipes in the ground coil

It will be observed that the effect of each successive term becomes increasingly less as the value of S increases. For the ground coil used in these tests the effect of the coils in an adjacent trench located 15 feet away amounts to only 5 per cent.

Having accounted approximately for the geometry, it is then required only to determine what empirical earth conductivity will make the Kennelly formula represent each particular rate of heat extraction which was observed during the tests. Since the empirical earth conductivity includes among other things the composite effects of moisture migration, earth porosity, averages of transient phenomena and freezing to the extent not otherwise accounted for, it should not be expected to equal the earth conductivity K associated with the Kennelly formula. Consequently it is given a different symbol E. The average rate of heat extraction by the earth coil then becomes

$$Q=EL(T_a-T_c)/G, \text{ average Btu per hour} \qquad (2)$$

where

E = the empirical earth conductivity, for which representative values are given in Table I
L = total feet of pipe in the ground coil
T_a = average ambient earth temperature
T_c = average coil temperature
G = the geometric factor of the ground coil installation, equation 1, whose value is 1.2 for the ground coils in these tests as described by Harlow and Klapper[1] and shown in cross section in Figure 1 of this paper

The very consistent agreement between representative values of E for comparable test periods is in itself partial confirmation of the steady-state assumption underlying the use of equation 2. The successively higher values for the respectively shorter test periods are thought to reflect principally the transient effects of freezing which characterize performance of the coils during the shorter periods of operation at the higher heat extraction rates.

It is apparent from the data in Tables IV-VII that the heat withdrawn from the ground during the 1949–50 heating cycle

Table VI. Thermocouple Readings at Lansdowne in Degrees Fahrenheit

Date	Left					Location—Inches from Pipe in Horizontal Plane								Right
	90	48	24	12	4	On Pipe	$7^{1}/_{2}$ (Between)	On Pipe	4	12	24	48	90	600
10/12/48	61.4	61.4	61.0	61.4	61.0	60.0	60.6	59.6	59.6	60.6	60.6	59.9	60.8	60
10/19/48	61.7	61.7	61.0	61.0	61.0	60.0	58.7	59.7	59.0	59.9	59.9	60.3	60.5	60
10/26/48	61.6	61.6	61.2	60.2	60.2	57.0	56.6	56.9	59.1	60.2	61.6	60.5	60.2	59
11/2/48	59.6	58.9	58.9	58.2	58.2	57.0	55.5	55.7	57.6	57.8	59.1	58.9	58.9	58.5
11/9/48	59.4	59.4	59.1	57.3	57.6	56.5	56.2	56.9	56.9	58.4	58.5	58.5	59.4	59.0
11/16/48	58.3	58.3	57.6	56.2	55.1	53.5	50.8	53.1	52.6	54.0	55.4	55.4	56.5	57.5
11/23/48	56.4	56.4	55.5	54.5	53.6	51.5	52.4	50.0	52.8	53.6	56.0	53.6	55.5	56.0
11/30/48	57.3	57.3	56.3	55.4	53.8	50.0	51.4	50.2	51.8	54.6	56.3	56.1	57.3	58.0
12/7/48	55.7	54.9	53.1	53.1	51.7	50 0	49.6	49.1	50.5	52.7	54.9	54.9	54.9	56.0
12/14/48	54.4	54.4	52.6	51.8	50.9	49.0	45.9	48.6	48.6	50.4	53.4	53.9	53.9	55.0
12/21/48	54.2	54.2	52.5	49.0	47.0	40.0	42.6	41.0	45.4	49.0	51.5	51.5	51.5	54.0
12/28/48	51.7	50.8	48.9	44.9	44.1	36.0	39.7	35.8	40.1	44.9	48.5	49.3	49.8	52.5
1/4/49	50.7	50.2	48.6	46.2	45.0	35.5	39.7	37.2	41.8	44.6	48.6	48.6	48.6	50.0
1/11/49	47.6	47.6	46.6	44.1	44.1	38.0	38.6	38.6	41.4	42.6	45.0	45.0	45.8	48.0
1/18/49	48.1	47.7	46.5	44.6	43.7	36.0	39.3	37.1	40.1	43.7	46.5	46.5	46.9	48.5
1/25/49	47.9	47.9	44.7	44.3	43.9	38.0	39.0	35.5	40.7	43.5	44.7	44.7	46.5	48.5
2/1/49	47.7	48.6	46.9	42.9	41.6	35.0	37.7	36.5	38.6	42.9	45.6	46.9	46.1	48.5
2/8/49	46.2	46.2	45.0	41.4	41.8	36.0	38.6	36.2	38.6	41.8	45.0	45.4	45.8	48.0
3/9/49					35.2	29.0	34.4	29.6	34.4					47.0
3/23/49			41.1	39.4	39.4	31.0	35.4	32.2	35.4	39.4	41.1			46.0
4/6/49			43.1	39.1	38.3	33.5	35.3	33.5	37.3	40.8	43.1			46.0
10/4/49	65.6	64.8	64.8	63.0	63.0	64.0	63.5	64.0	63.5	64.4	66.0	66.5	65.6	63.0
10/18/49	66.0	66.5	61.8	63.0	62.5	57.0	61.6	57.2	62.0	65.2	66.5	66.5	66.0	64.0
11/1/49	63.5	63.0	62.5	59.5	59.5	51.0	62.0	51.0	62.0	64.0	64.4	67.0	67.0	63.0
11/15/49	59.5	58.5	57.2	55.5	55.0	50.0	54.0	50.0	55.0	57.6	58.5	58.5	59.5	60.0
11/29/49	56.4	56.8	53.2	49.2	47.5	35.5	44.0	36.4	45.2	50.0	54.0	55.0	55.0	58.5
12/6/49	53.6	52.0	49.2	45.2	44.4	31.5	40.4	32.5	41.6	48.0	50.0	52.0	52.0	55.0
12/13/49	48.8	47.5	44.4	41.2	40.4	32.0	36.4	32.0	37.6	41.6	45.6	46.5	48.0	52.4
12/20/49	48.0	45.6	44.0	40.0	38.0	29.5	34.0	30.0	36.8	40.8	44.8	46.0	47.0	50.5
12/27/49	45.6	45.2	37.2	39.0	37.6	32.5	35.5	33.0	37.2	40.0	41.6	44.0	45.6	51.0
1/3/50	46.0	45.6	43.0	40.4	38.5	31.0	35.5	32.0	37.2	40.8	44.4	45.6	48.0	50.0
1/10/50	45.2	44.4	41.6	39.5	37.2	29.0	35.5	30.0	36.8	40.0	44.4	44.4	46.5	50.0
1/17/50	45.6	44.4	41.6	39.5	37.2		35.5	32.0	37.2	40.8	44.4	44.8	46.5	50.0
1/24/50	37.2	41.6	40.0	36.8	35.5	31.0	32.5	31.2	33.0	37.6	40.4	41.2	44.4	47.5
1/31/50	43.0	41.6	40.0	36.8	35.5	29.0	32.5	29.0	35.5	38.0	41.2	41.6	44.8	48.4
2/7/50	44.4	41.6	40.0	36.8	36.8	29.0	33.0	29.0	35.5	38.5	41.6	42.6	45.2	48.4
2/14/50	41.2	40.4	37.6	35.5	35.0	31.5	31.2	31.2	34.0	37.2	40.0	41.2	44.0	47.5
2/21/50	41.2	40.4	38.0	35.5	34.0	27.0	30.8	26.8	33.0	37.2	40.0	41.2	43.5	47.5
2/28/50	41.6	41.2	39.0	36.0	35.5	30.5	31.6	30.8	34.0	37.2	40.4	41.2	43.5	47.0
3/7/50	46.0	45.2	43.0	40.8	40.0	37.5	38.5	38.5	38.5	41.2	44.4	45.6	48.0	50.0
3/21/50	41.6	40.4	40.0	37.6	39.5	36.0	35.5	36.0	37.2	40.0	44.8	41.2	43.5	45.6
4/4/50	44.4	44.4	43.5	41.6	43.5	36.0	42.0	42.0	44.0	43.5	44.4	44.0	45.2	48.4

was completely restored before the start of the 1950–51 heating cycle. This should allay any apprehensions concerning the ability of the earth to recover completely between successive heating seasons.

Similar data were not taken during operation on the cooling cycle, with earth as a heat sink. Experience indicated that operation was entirely satisfactory. In this case, however, the earth moisture migrates away from the coil and reduces the apparent value of E. Measurements made on the soil at Whitemarsh[5] indicate a conductivity $K=0.8$ for heat dissipation, which suggests a corresponding summer season value of E equal to 0.9, or 1.0 for the earth as a heat sink.

Further analysis probably can contribute little practical increase in the accuracy of forecasting coil performance at comparable locations and under comparable operating conditions. A little closer examination of the freezing effects, however, may be instructive.

General Effects of Freezing

The preceding analysis has been based upon continuous chart records of thermocouple readings and heat extraction rates. There were, however, several exploratory thermocouples installed in the earth environment, 12 in the horizontal plane through the center line of the coil, and six in the vertical plane, midway between the two pipes in the trench, as shown in Figure 1. A single reading was made of each of the exploratory thermocouples and recorded during each weekly inspection of the installation. From these readings as given in Tables IV-VII, it is noted that the earth thermocouples 4 inches outward from the pipe indicate temperatures above freezing at all times during the two seasons, except at Whitemarsh where the temperature dropped to 31 degrees Fahrenheit on April 6, 1949, to 32 degrees Fahrenheit on February 21, 1950, and to 32 degrees Fahrenheit on March 8, 1950. The thermocouples located midway between the pipes in the coil plane indicated freezing temperatures occasionally at other times, but the instances cited were the only occasions when sufficient information was obtained for an approximate appraisal of the effect of freezing in terms of the generalized isotherms which have been plotted in Figure 2.

At the time of the readings on February 21, 1950, for instance, it is apparent that the "frost line" had just reached a distance 4 inches radially outward from the respective pipes in the horizontal plane. Since $S=15$ inches, $d_1=4$ inches, $d_2=19$ inches at 32 degrees Fahrenheit in this case, the corresponding isotherm is that for $(d_1d_2/S^2)=0.338$. By interpolation from Figure 2, it is found that the total enclosed cross section presumed to be frozen is $0.45S^2$ or 102 square inches. This represents 0.71 cubic feet of earth and 16.6 pounds of water per foot of pipe, if it is assumed that the 37.5-per-cent porosity of the earth is completely saturated with water. The corresponding latent heat of freezing is 2,400 Btu per foot of pipe, or equivalent to but 3 days' operation at the 35.0 Btu per hour average rate of heat extraction during the preceding week, The corresponding equivalent radius of the frost line is (d_1d_2/S) or 0.42 foot per pipe.

Two major changes accompany freezing. The principal effect is the release of heat from the immediate environment of the heat collecting pipe. A secondary effect is that the earth has about 50 per cent greater thermal conductivity after it is frozen. Both effects operate jointly to

Table VII. Thermocouple Readings at Lansdowne in Degrees Fahrenheit

	Location—Feet Below Ground Level											
	Center Line of Trench						Ambient Earth					
Date	2	3	4	5	6	7	2	3	4	5	6	7
10/12/48	59.4	59.9	60.6	60.6	59.7	59.7	59.7	59.9	59.2	60.0	61.0	59.9
10/19/48	58.5	58.8	58.8	58.7	58.1	58.8	57.8	59.9	60.3	60.0	60.2	61.4
10/26/48	56.6	58.4	58.4	56.6	59.4	59.6	57.6	59.4	59.4	59.0	60.8	62.0
11/2/48	55.1	56.0	55.7	55.5	57.5	57.8	56.0	58.2	57.5	58.5	59.3	59.3
1 /9/48	58.0	57.3	57.3	56.2	57.4	59.1	58.0	59.1	58.4	59.0	59.1	59.8
11/16/48	53.6	53.2	52.6	50.8	53.6	55.1	54.4	56.9	57.2	57.5	58.7	57.6
11/23/48	51.5	52.8	52.8	52.4	53.5	54.1	52.8	54.5	56.0	56.0	56.8	56.8
11/30/48	53.3	53.3	53.3	51.4	53.8	57.3	53.8	56.3	57.8	58.0	58.2	59.8
12/7/48	49.6	49.6	49.6	49.6	52.6	55.7	51.7	53.6	55.7	56.0	56.6	58.1
12/14/48	47.4	48.2	48.2	45.9	49.9	53.0	49.4	52.6	54.9	55.0	55.4	57.0
12/21/48	46.2	47.0	44.5	42.6	47.0	52.0	47.5	51.5	53.8	54.0	54.6	56.5
12/28/48	43.3	43.3	40.9	39.7	44.1	49.3	44.1	48.5	49.8	52.5	53.3	54.3
1/4/49	43.2	43.2	41.8	39.7	44.2	48.6	44.2	47.2	48.6	50.0	51.2	53.0
1/11/49	40.6	40.6	38.6	38.6	42.2	45.0	44.1	45.8	46.6	48.0	49.4	49.4
1/18/49	41.6	41.6	39.7	39.3	42.5	45.6	46.1	46.5	47.3	48.5	49.6	50.9
1/25/49	43.5	40.3	40.3	39.0	40.7	43.9	44.7	47.9	48.3	48.5	48.5	49.1
2/1/49	42.9	42.9	40.1	37.7	42.5	46.1	43.7	46.9	47.7	48.5	49.1	50.1
2/8/49	41.0	41.0	41.0	38.6	41.0	45.8	42.6	45.4	46.6	48.0	49.8	50.2
3/9/49				34.4						47.0		
3/23/49			36.6	35.4	36.2					46.0		
4/6/49			39.5	35.3	37.3					46.0		
10/4/49	64.4	66.0	64.0	63.5	63.5	64.0	61.6	62.0	62.5	63.0	62.5	63.0
10/18/49	67.0	66.0	63.5	61.6	63.0	64.0	63.5	63.5	63.5	64.	63.5	63.5
11/1/49	64.0	63.5	62.5	62.0	63.5	64.0	61.2	62.0	62.5	63.0	64.0	64.0
11/15/49	55.0	55.0	55.0	54.0	56.0	57.6	57.2	58.5	60.0	60.0	60.4	60.0
11/29/49	48.4	48.8	46.0	44.0	49.2	54.0	50.0	54.5	55.5	58.5	58.5	58.5
12/6/49	43.5	44.8	41.6	40.4	46.5	50.0	46.5	52.0	53.6	55.0	55.5	56.4
12/13/49	40.0	38.5	37.2	36.4	40.4	45.6	40.8	47.5	50.0	52.4	52.8	53.2
12/20/49	39.5	37.2	36.8	34.0	40.4	45.2	44.4	46.5	49.0	50.5	52.0	52.8
12/27/49	39.5	39.5	37.2	35.5	40.0	44.4	44.8	47.0	48.8	51.0	50.5	51.0
1/3/50	40.0	39.5	37.2	35.5	40.0	44.4	45.6	48.8	49.6	50.0	51.0	52.4
1/10/50	40.4	40.4	36.8	35.5	39.0	43.5	45.6	48.4	48.8	50.0	50.0	51.0
1/17/50	40.4	40.0	37.2	35.5	39.5	44.0	46.0	48.8	48.8	50.0	50.5	51.0
1/24/50	36.8	36.0	35.5	32.5	36.8	41.6	44.0	45.2	46.0	47.5	48.4	48.8
1/31/50	40.0	37.2	36.0	32.5	36.8	41.2	45.6	46.0	48.8	48.4	48.8	49.2
2/7/50	38.0	37.2	36.4	33.0	37.2	41.2	44.4	46.0	48.4	48.4	49.2	49.6
2/14/50	37.2	35.0	35.5	31.2	35.0	40.0	44.0	45.2	45.6	47.5	48.4	48.8
2/21/50	37.2	36.4	33.0	30.8	36.4	40.0	42.5	44.8	45.6	47.5	48.0	48.8
2/28/50	36.8	36.4	34.5	31.6	36.4	40.0	41.2	44.8	45.6	47.0	48.0	48.8
3/7/50	40.0	40.0	37.6	38.5	46.0	44.4	45.2	48.0	49.2	50.0	51.0	53.2
3/21/50	39.5	38.5	36.4	35.5	39.5	41.6	43.5	44.8	46.0	45.6	46.5	48.8
4/4/50	44.4	43.5	41.6	42.0	43.0	44.0	48.4	46.5	48.4	48.4	47.5	48.8

increase the heat extraction capacity of the ground coil. The portion of the earth within the frost line therefore carries the latent heat of fusion, as released at the frost line, plus the average of "base load" rate of heat transfer from ambient earth to the pipe. As the frost line progresses radially further from the pipe, the accompanying increase in latent heat transfer distance is offset to the extent that the increased proportion of higher conductivity frozen earth increases the effective conductivity for base load heat transfer from the ambient earth to the coil.

From a theoretical standpoint under steady-state conditions, therefore, the heat delivered to the coil per unit of temperature differential between coil and ambient earth should be a function of the extent of the frozen section. Furthermore the depression of the coil temperature below freezing per unit of heat withdrawn should be indicative of the extent of the frozen section.

Unfortunately, because the rate of heat withdrawal was continually fluctuating, it has been impossible to evaluate these effects quantitatively as a basis for predicting heat withdrawal rates under other differing modes of operation. Certain general conclusions have been drawn, however, from an analysis of the data obtained.

The average rate of heat extraction per degree Fahrenheit prior to freezing was about 1.1 (Btu)/(hour) (degrees Fahrenheit) (foot) which corresponds to a value of $E=1.3$ for use in equation 2. When the coil temperature was depressed below 32 degrees Fahrenheit, the average rates of heat extraction per degree Fahrenheit increased and became fairly constant. The average rates so observed varied from 1.3 to 1.8 corresponding to values of E of 1.55 and 2.2 respectively. The highest rate observed on a daily basis was 2.1 corresponding to a value of 2.5 for E.

The increased rates of heat extraction during frozen operation are not attributable to latent heat of freezing alone. For example consider the situation existing on February 21, 1950 where the frozen section was known to have built up to an equivalent radius of 0.42 foot in a matter of 3 weeks, liberating a total of 2,400 Btu per foot due to latent heat of freezing, at an average rate of 4.7 (Btu)/(hour) (foot) during the 3 weeks. The corresponding G for the preceding equivalent radius is 0.92. The average heat withdrawal rate on February 21, 1950 was 39.8 (Btu)/(hour) (foot), the ambient earth temperature was 47 degrees Fahrenheit, and the coil temperature was 28 degrees Fahrenheit.

Taking the value of $E=1.3$ as already derived for unfrozen operation, and substituting the preceding values in equation 2 gives

$$Q/L=\frac{1.3(47-32)}{.92}=21.2\text{ (Btu)/(hour)(foot)}$$

as the expected heat withdrawal rate in the unfrozen earth. The net value of 39.8—4.7=35.1 observed is nearby 70 per cent higher than the preceding value of 21.2. Moisture migration is thought to be the only supplementary mode of heat transfer which might account for the additional heat transfer effect.

Moisture Migration

Whenever the air spaces in the earth environment are only partly saturated with moisture, the introduction of a temperature differential causes a redistribution by migration of the moisture toward the region of lowest temperature. Stability is reached when the unbalanced forces of surface tension, vapor pressure, and gravity reach equilibrium. If there is sufficiently high water saturation at the cold surface, the conditions of stability may include some continual loss of moisture by gravitational drainage from the water saturated region. Such drainage would require the continuous supply of moisture from the warm regions toward the cold surface. This mode of moisture and heat transfer would have the effect of increasing the apparent conductivity of earth in well-drained ground coil locations and probably is related to the porosity of the earth. For such locations it is possible that the use of a coarse backfill would be preferable to clay, although the effects of freezing and thawing may produce a degree of porosity in clay and drainage comparable to that obtained by the use of sand or crushed stone.

It is, of course, possible for moisture to migrate in either the liquid or the vapor phase. The available information[4,5] on moisture migration in soil due to thermal gradients (sometimes called thermal osmosis), however, indicates that liquid phase migration predominates. To account for the net 35.1—21.2=14 (Btu)/(hour) (foot) which might be attributed to moisture migration from ambient earth at 47 degrees Fahrenheit to 32 degrees Fahrenheit at the frost line would require the continuous movement of about 3/4 pint of water or 40 cubic feet of water vapor per hour and foot of pipe. This seems almost reasonable, but rather larger than expected. The natural breathing of the earth pores, with day-to-day variations in barometric pressure, however, might cause substantially more vapor flow than would be induced by the temperature gradient alone. Also the rain and snow fall would be expected to supplement the thermal osmosis effect.

The theoretical calculations before freezing previously mentioned indicate an earth thermal conductivity $K=1.0$ Btu/(hour) (degrees Fahrenheit) (foot) at Whitemarsh which is slightly higher than the value of 0.9 as calculated for this soil completely saturated with water.[3] It would appear therefore that a certain amount of continuous moisture migration was also present at coil temperatures above freezing.

The possibility of moisture migration, therefore, must be kept in mind as an explanation of the comparatively high heat extraction rates observed. It must be realized also, that if the frozen section is extended appreciably beyond that observed either by longer periods of frozen operation or by decreasing the coil temperature below 27 degrees Fahrenheit, the conditions governing this effect may be altered and the observed rates of heat withdrawal may not be maintained. This suggests that values of the empirical constant E which were representative of the maximum rates observed, may overstate the ground coil output to be expected for prolonged operation at a uniformly high extraction rate or at coil temperatures below 27 degrees Fahrenheit.

Conclusions

1. Earth is an entirely practical and reliable source of heat for the heat pump. It is an equally practical heat sink, but has a lower heat transfer rate, requiring a longer coil for equal transfer.

2. The heat withdrawn is completely restored by the end of an annual cycle, thus permitting the continuing use of a system without loss of efficiency.

3. The rate of heat transfer depends essentially upon the geometry of the coil and the temperature differential between coil and ambient earth. The highest heat transfer rates were obtained when the temperature of the coil was depressed below freezing. The maximum rate of heat withdrawal observed was about 40 Btu/(hour)(foot) with a temperature differential of 20 degrees Fahrenheit and a coil temperature of 27 degrees Fahrenheit.

References

1. Residential Heat Pump Experiments in Philadelphia—Installation and Operating Experience, **J. H. Harlow, G. E. Klapper.** *AIEE Transactions*, volume 71, part II, 1952 (January 1953 Section) (*Paper 52-342*).

2. The Temperature Rise of Buried Cables and Pipes, **J. H. Neher.** *AIEE Transactions*, volume 68, part I, 1949, pages 9–17.

3. The Thermal Conductivity of Moist Soil, **A. S. Mickley.** *AIEE Transactions*, volume 70, part II, 1951, pages 1789–94.

4. The Thermal Movement of Moisture in Soil, **A. S. Mickley.** *AIEE Transactions*, volume 68, part I, 1949, pages 330-33.

5. Fundamental Similarities Between Electro-Osmotic and Thermo-Osmotic Phenomena, **H. F. Winterkorn.** *Proceedings*, Highway Research Board N.D.R.C. (Washington, D. C.), 1947, pages 443–55.

No Discussion

Measurement of Cathodic Protection Currents from Sacrificial Anodes

H. N. HAYWARD
ASSOCIATE MEMBER AIEE

R. M. WAINWRIGHT
MEMBER AIEE

THE measurement of currents and voltages in low-energy circuits always presents certain difficulties. The presence of the measuring equipment causes changes in the circuit parameters, and even though an exceedingly accurate instrument is used, the instrument reading may be far from the true value obtained when the instrument is not present.

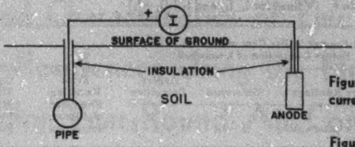
In cathodic protection work use is made of artificial galvanic circuits set up by the metal structure to be protected as a cathode, a sacrificial material such as magnesium or zinc as an anode, and the soil or other aqueous solutions as the electrolyte. The circuits involved may be represented for the simplest case as in Figure 1(A), and from an equivalent circuit standpoint as in Figure 1(B). The symbols used in these figures are defined as follows:

E_{12} = the metallic surface potential-to-soil
E_{34} = the anode potential-to-soil
R = the total circuit resistance including the soil and the connecting wire between the anode and the metallic structure, but not including the ammeter
R_{I_1} = the resistance added to the circuit by the ammeter

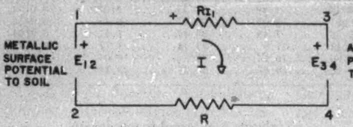
If the ammeter resistance R_{I_1} were zero, a certain current, I, would flow in the circuit. However, if the ammeter resistance is not zero, the current would drop to some new value, I_1, which is now measured by the instrument. In no case will the instrument ever measure the correct value I, except that if R_{I_1} is much less than R, $I=I_1$ approximately.

Paper 52-305, recommended by the AIEE Chemical, Electrochemical and Electrothermal Applications Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted June 12, 1952; made available for printing September 26, 1952.

H. N. Hayward and R. M. Wainwright are with the University of Illinois, Urbana, Ill.



(a) ACTUAL CIRCUIT



(b) EQUIVALENT CIRCUIT

Figure 1 (left). The basic current measuring circuit

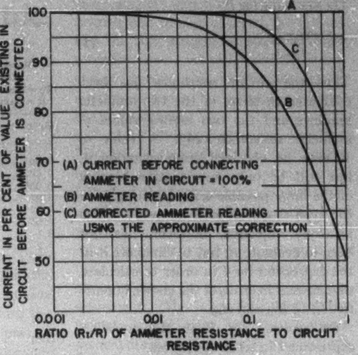


Figure 3 (right). Effect of ratio of ammeter resistance to circuit resistance on reading of ammeter

In practice the currents from single anodes may be in the normal range of from 40 to 400 milliamperes, with values of R in the range of 2 ohms to 40 ohms. Milliammeters which usually would be available to measure such anode currents may have internal resistances from 0.01 ohm to as high as 2 ohms. These ranges of values indicate that ratios of R_I/R may be obtained from 1/4000 to 1/1. If the ratio is between 1/4000 and 1/100, the milliammeter may be used with less than 1 per cent error due to this factor. However, if the ratio lies between 1/100 and 1/1, the error becomes an important item. In the extreme case the error may be as much as 50 per cent or more in terms of the correct value of current. Since engineers rarely know, or have the time and equipment to determine, the range of ratios that exists in a particular situation, the errors in reading anode currents frequently become a factor of great uncertainty.

To solve this error problem, engineers have resorted to estimating the errors, to the use of "zero-resistance" ammeters, or to the use of permanent shunts installed in series with the anodes. The first method is obviously not desirable. The second involves the introduction of an additional voltage in series with the ammeter circuit which compensates for the voltage developed across the internal resistance of the instrument. This zero-resistance method gives adequate results, but the instrument is more expensive and the process of obtaining a reading is similar to that for a potentiometer. Such ammeters are more suited for laboratory use than for field use as a rule. The permanent shunt is also satisfactory in many cases, but frequently the added resistance in the circuit is undesirable, and the cost and maintenance of the shunt is always a factor. Calibration problems and the necessity of using a special set of leads and ammeter also create disadvantages.

The authors have made an attempt to outline an additional solution to the problem through the use of a "double-reading" taken from a comparatively sturdy and inexpensive instrument. The method is capable of very exact results, and no self-contained batteries or elaborate circuit arrangements are necessary. The analysis of such an instrument is built up as follows:

Refer to Figure 1(B): when the ammeter is not in the circuit, current will be

$$I=\frac{E_{12}-E_{34}}{R}, \; (R_{I_1}=0)$$

and is the value desired when making a measurement. But adding the ammeter resistance to the circuit increases the circuit resistance by the value R_{I_1}, so that the measured value of current is now

$$I_1=(E_{12}-E_{34})/(R+R_{I_1})$$

The ratio of the initial and the measured currents is then

$$\frac{I}{I_1}=\frac{\dfrac{E_{12}-E_{34}}{R}}{\dfrac{E_{12}-E_{34}}{R+R_{I_1}}}=\frac{R+R_{I_1}}{R}$$

That is, the ratio of the two values of current is the inverse of the ratio of the two corresponding values of the circuit resistance. I may be expressed as

$$I=I_1\left(\frac{R+R_{I_1}}{R}\right)$$

but cannot be determined because the circuit resistance R is not known.

If a second ammeter, or the same one modified, of higher known resistance, R_{I_2}, is connected in the circuit, the new measured value of current will be

$$I_2=(E_{12}-E_{34})/(R+R_{I_2})$$

Again the ratio of the initial current and this second measured value is

$$\frac{I}{I_2}=\frac{\dfrac{E_{12}-E_{34}}{R}}{\dfrac{E_{12}-E_{34}}{R+R_{I_2}}}=\frac{R+R_{I_2}}{R}$$

and I may be expressed as

$$I=I_2\left(\frac{R+R_{I_2}}{R}\right)$$

Combining these two expressions for I gives

$$I=I_1\left(\frac{R+R_{I_1}}{R}\right)=I_2\left(\frac{R+R_{I_2}}{R}\right)$$

or $I_1(R+R_{I_1})=I_2(R+R_{I_2})$

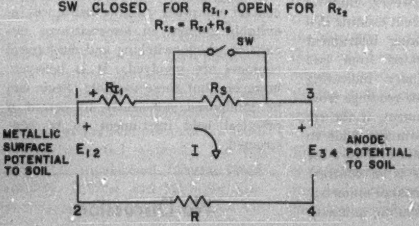


Figure 2. Measuring circuit modified to increase the ammeter resistance

Then, solving for R

$$R=\frac{I_2R_{I_2}-I_1R_{I_1}}{I_1-I_2} \qquad (1)$$

Thus the circuit resistance has been evaluated in terms of the two ammeter readings and the two known values of ammeter resistance only. A value of R may be obtained from equation 1, but will be only approximate because of the large possible error resulting from the type of calculations necessary. The principle objective, however, has not been to evaluate R, but to eliminate it in the expression for I in order to calculate that current in terms of the known quantities I_1, I_2, R_{I_1}, and R_{I_2}.

Substituting for R in the equation expressing I in terms of I_1 gives

$$I=I_1\left(\frac{R+R_{I_1}}{R}\right)=I_1\left[\frac{\left(\dfrac{I_2R_{I_2}-I_1R_{I_1}}{I_1-I_2}\right)+R_{I_1}}{\dfrac{I_2R_{I_2}-I_1R_{I_1}}{I_1-I_2}}\right]$$

or, simplifying

$$I=I_1\left[\frac{I_2(R_{I_2}-R_{I_1})}{I_2R_{I_2}-I_1R_{I_1}}\right]$$

If the same ammeter is used for both measurements, $R_{I_2}=R_{I_1}+R_s$, as shown in Figure 2. Then

$$I=I_1\left(\frac{I_2[(R_{I_1}+R_s)-R_{I_1}]}{I_2(R_{I_1}+R_s)-I_1R_{I_1}}\right)$$

$$=I_1\left(\frac{I_2R_s}{I_2R_s-R_{I_1}(I_1-I_2)}\right) \qquad (2)$$

This equation shows how the current existing in the circuit before an ammeter is connected may be calculated in terms of the two ammeter readings and the known values of resistance added to the circuit by connecting in the ammeters. But the calculation requires several operations, hence it is not practical for general use.

It should be noted that as the resistance added to the circuit is increased, the current decreases. If a relationship between the two values of ammeter resistance can be established so that $(I-I_1)=(I_1-I_2)$, or $I=I_1+(I_1-I_2)$, the difference between the two readings added to the first reading will give the true value of current existing in the circuit before the connection of either of the instruments. From

$$I=I_1\left(\frac{R+R_{I_1}}{R}\right) \quad \text{and} \quad I=I_2\left(\frac{R+R_{I_2}}{R}\right)$$

it follows that

$$\frac{I_2\left(\dfrac{R+R_{I_2}}{R}\right)}{I_1\left(\dfrac{R+R_{I_1}}{R}\right)}=1$$

Table I. Values for I_1, I_2, and $I_1+(I_1-I_2)$

Resistance Ratios R_I = Ammeter Res. R = Circuit Res.		Measured Current in Per Cent of Current in Circuit Before Connection of Ammeter				Per Cent Error Resulting From Use of Approx. Correction	Ratio of Current Diff. to Higher Current Reading
		Ammeter Readings		Difference	Approx. Correction		
R_I/R	R/R_I	I_1 (R_{I_1})	I_2 ($R_{I_2}=2R_{I_1}$)	(I_1-I_2)	$I_1+(I_1-I_2)$		
0.001	1,000	99.9	99.8	0.1	100	0	0.0010
0.002	500	99.8	99.6	0.2	100	0	0.0020
0.005	200	99.5	99.01	0.49	99.99	0.01	0.0049
0.01	100	99.01	98.04	0.97	99.98	0.03	0.0098
0.02	50	98.04	96.15	1.89	99.93	0.07	0.0193
0.05	20	95.24	90.91	4.33	99.57	0.43	0.0455
0.1	10	90.91	83.33	7.48	98.49	1.51	0.0823
0.2	5	83.33	71.43	11.9	95.23	4.77	0.143
0.5	2	66.67	50.00	16.67	83.34	16.66	0.250
1.0	1	50.00	33.33	16.67	66.67	33.33	0.333

then

$$\frac{I_2}{I_1}=\frac{R+R_{I_1}}{R+R_{I_2}} \quad \text{or} \quad I_2=I_1\left(\frac{R+R_{I_1}}{R+R_{I_2}}\right)$$

Then $(I-I_1)=(I_1-I_2)$ becomes

$$I_1\left(\frac{R+R_{I_1}}{R}\right)-I_1=I_1-I_1\left(\frac{R+R_{I_1}}{R+R_{I_2}}\right)$$

Dividing through by I_1 and rearranging gives

$$R_{I_1}(R+R_{I_2})=R(R_{I_2}-R_{I_1})$$

and solving for R_{I_2} gives

$$R_{I_2}=\frac{2RR_{I_1}}{R-R_{I_1}} \qquad (3)$$

As R_{I_1} and R_{I_2} usually be will considerably less in value than R, it can be seen that R_{I_2} must be slightly more than twice R_{I_1}, depending on the value of R, for the relation $I=I_1+(I_1-I_2)$ to hold. But R is variable and generally is not known; hence it is impossible to obtain the desired correction exactly by using fixed values of R_{I_1} and R_{I_2}. However, if $R_{I_2}=2R_{I_1}=R_{I_1}+R_s$, where $R_s=R_{I_1}$, equation 2 becomes

$$I=I_1\left[\frac{I_2}{I_2-(I_1-I_2)}\right] \qquad (4)$$

Although the expression now contains only the two measured values of current, each calculation for making a correction involves two subtractions, a division, and a product, a total of four steps. It should be noted, however, that by this method the true current existing in a circuit with no measuring instrument present can be calculated from two readings of two ordinary indicating instruments, or from two readings from a single modified instrument, if the resistances which the instrument adds to the circuit are in the ratio of 2 to 1 for the two readings. This is in contrast to the use of the zero-resistance ammeter, or to the practice of installing ammeter shunts as referred to previously.

For many purposes an approximate correction can be made that is much simpler than equation 4, yet this correction will give the value of I (the true current with no ammeter in the circuit) to within 0.5 per cent, added to the uncertainty in the measured values themselves, provided that the ratio of ammeter resistance to circuit resistance is not more than 0.05. Then

$$I=I_1+(I_1-I_2), \text{ approximately} \qquad (5)$$

Values for I_1 and I_2 (for the condition $R_{I_2}=2R_{I_1}$), and values for $I_1+(I_1-I_2)$ are tabulated in Table I. It can be seen that the sum of $I_1+(I_1-I_2)$ is within 0.43 per cent of the true value of I for all values of the ratio of ammeter resistance to circuit resistance less than 0.05. This is in contrast to a difference of nearly 5 per cent between an uncorrected ammeter reading, I_1, and the true current I, providing that the ammeter resistance is not more than one-twentieth the circuit resistance. Figure 3 shows some of the results of Table I in graphic form.

An ammeter incorporating the principles shown in Figure 2 to permit two known values of resistance to be inserted in the circuit is being constructed and will be used for making field measurements in the near future. Such an instrument can be built to provide this double resistance feature on each of several ranges. Some inaccuracy will result because of the relative importance of contact and other variable resistances inevitably present in low-resistance circuits in which switching and mechanical contacts are involved. It is believed, however, that these variable effects can be minimized sufficiently to permit a practical field instrument to be constructed.

No Discussion

Load Characteristics of Five All-Electric Residences Using the Heat Pump for Year-Round Air Conditioning

PHILIP SPORN
FELLOW AIEE

E. R. AMBROSE
NONMEMBER AIEE

THE residential heat pump offers a tremendous potential load to the electrical utility industry. A load which, with any reasonable degree of public acceptance, easily can equal and perhaps exceed several times the present-day total domestic electrical consumption. The total estimated residential electrical consumption in the United States was 29.7 billion kilowatt-hours for 1944 and 67 billion kilowatt-hours for 1950. The last known residential heating installation survey, made in 1940, gave the total useful heating fuel consumption as $2,382 \times 10^{12}$ Btu which represents 697 billion kilowatt-hours when converted to equivalent electrical units and to approximately 155 billion kilowatt-hours if the heating was supplied by heat pumps having a coefficient of performance of 4.5.[1] Using these figures, a 20-per-cent heat pump saturation of residential heating installations would bring about a consumption equal approximately to the total residential electrical consumption in the United States, providing it can be assumed that the increase in the heating installations and in the residential electrical consumption since 1940 were in about the same proportion. In any case, this load may be of such magnitude that its effect on the present and future electrical utility systems needs to be evaluated fully. All possible information must be obtained on the load factor, on the probable demand coincident with the system's peak, on the diversity of a number of installations in any given area, and on the extent and duration of the peaks of the nondiversified demand in order to determine power cost as well as to provide suitable data for substation and feeder design and to permit proper sizing of distribution transformers.

Some useful data have been published on the electrical characteristics of heat pump systems in office buildings and other commercial establishments, but very little actual field test data are available on residential systems. For this reason the Appalachian Electric Power Company, Indiana and Michigan Electric Company, Kentucky and West Virginia Power Company, Inc., The Ohio Power Company, and Wheeling Electric Company, all part of the American Gas and Electric system, initiated during the latter part of 1948 and the early part of 1949 a field test program on five residential heat pump installations. The five installations, located in Kingsport, Tenn., Abingdon, Va., Roanoke, Va., Coshocton, Ohio, and South Bend, Ind., were made in newly constructed houses in co-operation with the home owners, their architect, and the manufacturer.[2,3] The units were located in five different geographic areas so that their performance could be observed under a wide variation of climatic conditions. In order to obtain some information on the relative merits of the various heat sources the Kingsport, Roanoke, and Coshocton systems were designed to use air exclusively as the heat source; South Bend air and the earth, and Abingdon air and well water.

Table I gives some of the most important design data; design details of the duct systems and of the structure for the five installations have been described previously.[2,3] The size, type, and the heating and cooling requirements of the houses vary considerably. The five different types of houses were included in order to get some data on the effect of the size and shape of a structure on the heating and cooling requirements. A self-contained unit, Figure 1, 78 inches long by 60 inches high by 40 inches deep, was installed in each of the five installations. Each unit has two refrigerant circuits consisting of a compressor, heating and cooling surface, valves, and connecting piping. The heating and cooling requirements for the conditioned space is produced by changing the direction of the refrigerant flow by means of a 4-way refrigerant valve.[4] The system furnishes year-round air conditioning and is completely automatic in operation.

The testing instruments, shown in Figure 2, were installed on each of the five installations to obtain all possible operating and performance data during the field test period. Pressure and temperature recorders were used to determine the heating and cooling effect of the system, and the volt, ampere, and kilowatt-hour recorders were used to obtain the electric energy consumption of the heat pump equipment and of the general appliances in the residences.

A 15-kw electric resistance heater was installed in each of the five heat pump units to serve as a stand-by in case of failure of the heat pump equipment and as a booster when the capacity of the heat pump is exceeded. Also, one day each week the electric resistance heaters were used exclusively to heat the residence in order to check the calculated heating requirement. The heat equivalent of the electric energy used by the resistance heaters during each of the 24-hour periods was plotted against the outdoor temperature for each of the installations. Figure 3 shows these data as well as the calculated heat loss for the Abindgon installation based on maintaining a 70-degree Fahrenheit indoor temperature. Each point on the figure represents a 24-hour period. The actual heat loss, based on the electric energy consumption of the resistance heater, was slightly higher than that calculated on all of the installations. It also was noticed that the heat loss at the same outdoor temperature varied considerably during the different test periods. Some of this variation is probably due to the changes in wind velocity, to the different amount of heating effect obtained from the sun, to the inconsistency of the living habits of the occupants, and to the fact that higher indoor temperature was maintained than the 70 degrees Fahrenheit used as a basis for the design calculations.

The expected heating output at various outdoor temperatures of one of the packaged heat pump units, using air as the heat source, is given in Figure 4. The calculated heat losses of the five installations, with a 70-degree Fahrenheit indoor temperature, also are included in this figure. It can be seen that the unit satisfies the calculated heating requirement at some point below 0 degree Fahrenheit for the Roanoke system, at approximately 0 degree Fahrenheit for Abingdon, 3 degrees Fahrenheit for Kingport, 7 degrees Fahrenheit for South Bend, and 16 degrees Fahrenheit for Coshocton. Since the

---

Paper 52-339, recommended by the AIEE Domestic and Commercial Applications Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted January 31, 1951; made available for printing September 3, 1952.

PHILIP SPORN and E. R. AMBROSE are both with the American Gas and Electric Service Corporation, New York, N. Y.

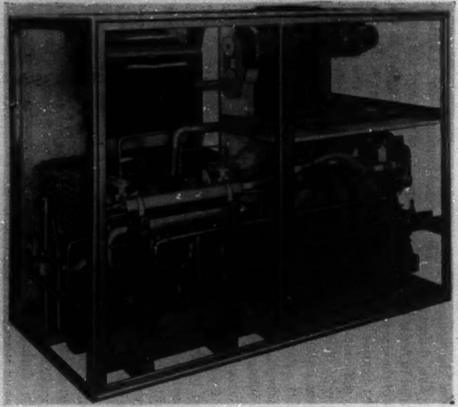
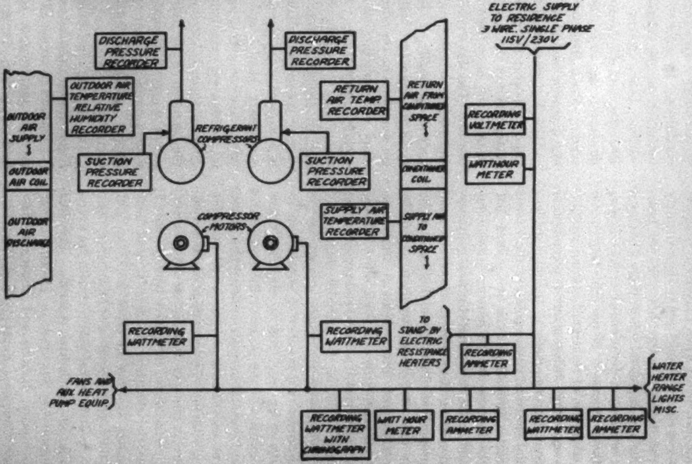


Figure 1 (left). The self-contained heat pump unit installed in each of the five residences with the front and sides removed

Figure 2 (below). Typical schematic diagram showing testing equipment for residential heat pump units


ELECTRIC SUPPLY TO RESIDENCE 3 WIRE. SINGLE PHASE 115V/230V
DISCHARGE PRESSURE RECORDER
DISCHARGE PRESSURE RECORDER
OUTDOOR AIR SUPPLY
OUTDOOR AIR COIL
OUTDOOR AIR DISCHARGE
OUTDOOR AIR TEMPERATURE RELATIVE HUMIDITY RECORDER
REFRIGERANT COMPRESSORS
SUCTION PRESSURE RECORDER
SUCTION PRESSURE RECORDER
RETURN AIR TEMP RECORDER
RETURN AIR FROM CONDITIONED SPACE
CONDITIONER COIL
SUPPLY AIR TO CONDITIONED SPACE
SUPPLY AIR TEMPERATURE RECORDER
RECORDING VOLTMETER
WATTHOUR METER
COMPRESSOR MOTORS
RECORDING WATTMETER
RECORDING WATTMETER
TO STAND-BY ELECTRIC RESISTANCE HEATERS
RECORDING AMMETER
FANS AND AUX HEAT PUMP EQUIP.
RECORDING WATTMETER WITH CHRONOGRAPH
WATT HOUR METER
RECORDING AMMETER
RECORDING WATTMETER
RECORDING AMMETER
WATER HEATER RANGE LIGHTS MISC.

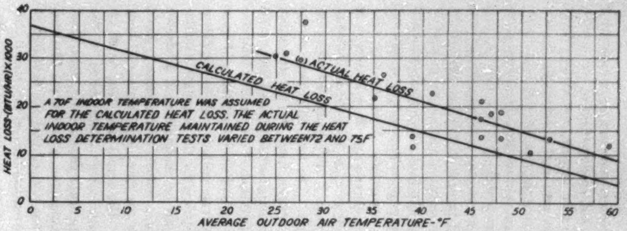
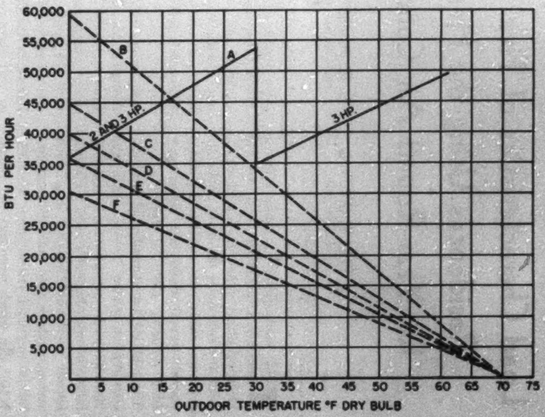


Figure 3 (above). The calculated heat loss and the actual heat loss, as determined by field test at various outdoor air temperatures for the Abingdon, Va., residence

Figure 4 (below). Expected heating output at various outdoor temperatures of package heat pump unit using air as heat source

A: Heating output of heat pump unit
B: Calculated heat loss, residence—Coshocton, Ohio
C: Calculated heat loss, residence—South Bend, Ind.
D: Calculated heat loss, residence—Kingsport, Tenn.
E: Calculated heat loss, residence—Abingdon, Va.
F: Calculated heat loss, residence—Roanoke, Va.



outdoor design temperature normally used for equipment selection is 0 degrees Fahrenheit for South Bend and Coshocton, and 10 degrees Fahrenheit for Kingsport, Abingdon, and Roanoke, it is evident that the Coshocton and South Bend units are somewhat undersized, and the Kingsport, Abingdon, and Roanoke are slightly oversized at their respective outdoor design temperatures. It is important to note that no allowances were made in the calculated heat loss for duct losses, for quick morning pickup, and for unusually high wind velocities. Also, no credit was made in the heat loss calculations for any useful solar heating effect which may be obtained. These installations purposely were made with a rather close margin between the capacity of the units and the heat loss of the structure to get some useful data for establishing sound design limits even though the procedure may be somewhat inconsistent with the general practice of sizing the heating plant to have a capacity of 50 to 100 per cent over the calculated heat loss. A great deal depends in this regard on how the system is operated normally and particularly on what inside temperature level is maintained during the night cycle. During the field test periods, the occupants of the five residences were permitted to raise or lower the inside temperatures to satisfy their individual comfort. The setting of the thermostats, however, never was changed purposely to maintain a lower night temperature.

Table II gives the degree days by months for each of the five locations. These data can be used in comparing the heating requirement of the five residences and in developing a heating factor which will be applicable to similar installations.

The cooking, water heating, and all other appliances are completely electric in the five residences. A list of the appliances installed (exclusive of the heat pump), and the connected load of each, is given in Table III. The connected load ranges from approximately 17.5 kw at Roanoke to 27.6 kw at South Bend. The average connected load for the five installations without the heat pump is about 21.0 kw.

The heat pump connected load in watts for the five installations, during the heating cycle, at an outdoor temperature of 18 degrees Fahrenheit, is given in Table IV. It is to be noted that the compressor load is quite variable with changes in outdoor temperature. The connected load given in Table IV increases with outdoor temperature until 30 degrees Fahrenheit is reached. Above 30 degrees Fahrenheit the compressor load is reduced ap-

Table I. Summary of Data on Five Pilot Residential Heat Pump Installations

Location	Kingsport, Tenn.	Abingdon, Va.	Roanoke, Va.	Coshocton, Ohio	South Bend, Ind.
No. of stories in house	1½ story	1 story	2 story with basement	1 story with basement	1 story
Volume, cubic feet*	20,710	14,670	13,450†	19,114†	11,500
Heat loss, Btu per hour‡					
Conduction	29,650	28,336	24,681	47,968	38,808
Infiltration	10,350	8,540	6,200	11,300	5,860
Total	40,000	36,876	30,881	59,268	44,668
Heat gain, Btu per hour‡					
Conduction and sun effect	30,666	19,262	14,466	14,241	19,949
Infiltration	4,260	3,515	2,550	4,680	2,405
Total	34,926	22,777	17,016	18,921	22,345
Ratio, window area to total wall area, per cent	13.1	12.3	13.1	30.5	21.0
Heat loss, Btu per hour per 1,000 cubic feet	1,940	2,510	2,300	2,730	3,880
Total number of degree days in normal heating season	3,885	3,980	4,075	5,535	6,202

* As represented by the outside dimensions for the area and the underside of the floor joist to the top of the ceiling as the height.

† Basement not included.

‡ A 70-degree Fahrenheit indoor-outdoor temperature differential was used in the heat loss calculations mainly to show what the relative heating requirements would be at the same design temperature. This figure is not intended as the design temperature to be used at the different locations. The outdoor design temperature normally used as a basis for the selections of equipment is 0 degrees Fahrenheit for South Bend and Coshocton, and 10 degrees Fahrenheit for Kingsport, Abingdon, and Roanoke.

The heat gain figures given are based on a 95 degree Fahrenheit dry bulb and 75-degree Fahrenheit wet bulb outdoor temperature and an 80-degree Fahrenheit dry bulb and 67-degree Fahrenheit wet bulb indoor temperature.

Table II. The Monthly Degree-Days for the Five Locations as Determined from the Average Daily Outdoor Temperature Recorded During the Test Period

	Kingsport, Tenn.	Abingdon, Va.	Roanoke, Va.	Coshocton, Ohio	South Bend, Ind.
May 1949	61	107	87	159	216
June			1	17	25
July					1
August		3		8	7
September	73	132	93	172	199
October	16	197	140	218	313
November	588	648	536	653	769
December	673	677	764	841	1,002
January 1950	470	531	534	733	1,040
February	557	773	686	858	1,103
March	605	670	687	852	1,035
April	351	444	346	544	712
Total	3,394	4,182	3,874	5,055	6,422

Table III. Connected Load of the General Appliances in Watts for the Five Residences

Appliances	Kingsport, Tenn.	Abingdon, Va.	Roanoke, Va.	Coshocton, Ohio	South Bend Ind.
Water heater	2,500	4,000	2,500	3,000	3,250
Range	8,400	9,900	11,000	10,500	11,300
Refrigerator	200	360	75	95	322
Electric dryer					4,700
Washing machine		1,170			400
Lights, radio, television	2,670	1,260	1,816	3,325	1,200
Fan	40			65	80
Vacuum cleaner	275			300	632
Iron	1,000	1,000	1,000	1,000	1,000
Miscellaneous:					
Toaster	1,150	660	1,200	1,230	12
Mixmaster	100				
Clock	2				
Sewing machine	100				
Heater	600			2,000	
Disposal				185	560
Electric blanket					207
Tea kettle					1,000
Ironer					1,680
Roaster					1,320
Home freezer		550			
Electric breeder		100			
Deepwell pump		2,100			
Total	17,037	21,100	17,591	21,700	27,663

Table IV. Heat Pump Connected Load in Watts for the Five Installations During the Heating Cycle at an Outdoor Temperature of 18 Degrees Fahrenheit

	Kingsport, Tenn.	Abingdon, Va.	Roanoke, Va.	Coshocton, Ohio	South Bend Ind.
Compressor number 1	2,330	2,330	2,330	2,330	2,330
Compressor number 2	1,940	2,330	1,940	1,940	1,940
Outdoor fan	715	715	715	715	715
Conditioner fan	900	900	900	900	900
Other		Deepwell Pump 2,200*		Circulating Pump 185†	Circulating Pump 250
Total	5,885	8,475	5,885	6,070	6,135

* Deepwell Pump used for domestic water supply and heat pump.

† Heat storage system incorporated in December 1949.

proximately 1,940 watts, since the 2-horsepower compressor is not operated at the higher outdoor temperatures.

During the test period, a complete record was kept of the electric energy consumption of the heat pump and the other appliances. These data are shown in Tables V, VI, and VII, and in Figures 5 through 8. Table V gives the monthly kilowatt-hours for the general electric appliances and for the heat pump during the heating and cooling cycle. The annual appliance load, less the heat pump, varies from 3,246 to 10,803 kilowatt-hours with an average for the five installations of about 7,600 kilowatt-hours. The annual heat pump load varies from 13,365 to 26,231 kilowatt-hours with an average for the five installations of about 20,000 kilowatt-hours. The annual total electric load for the five installations ranges from 16,611 to 33,824 kilowatt-hours with an average of about 27,000 kilowatt-hours. The average figures show that annual energy consumption of an all-electric residence with a heat pump is approximately $3^1/_2$ times as great as for the same residence without a heat pump.

Table VI gives the average 30-minute integrated demand and the monthly load factors for the five residences. The combined demand as well as the separate demands of the general appliances and the heat pump are given to show the relative effect of the various loads. These demands were used to calculate the monthly and yearly average load factors given in the last item of the table. It can be seen from Table VI that the general appliance yearly average load factor of the five residences varies from 9.6 per cent for the Roanoke to 19.5 per cent for the Coshocton installation with an average for the four installations of 13.4 per cent. The annual load factor for the combined load varies from 21.4 to 37.7 per cent with an average of 27.8 per cent for the four installations. The Kingsport installation was not included in this comparison because the system was found to be not operating properly and because it was felt that excessive heat losses were being experienced from the duct system and from the conditioned space. The data given in Table VI show that the heat pump system practically doubles the yearly average load factor in each case and materially improves the monthly load factor even during the summer period.

The daily load factor of a residence with the heat pump included may be much more important than the yearly average load factor when its effect on the distribution system is considered. The demand, kilowatt-hours, and load factor on a cold, a medium, and a warm day for four of the five installations are given in Table VII. Figures 5 through 8 show the daily kilowatt demand of the general applicances and of the combined load for four of the installations on a day ex-

Table V. The General Appliance and Heat Pump Monthly Kilowatt-Hours for the Five Residences

	1949								1950				
	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	March	April	Total
Kingsport, Tenn.													
Heating	444				387	658	3,715	3,560	2,915	2,955	2,983	1,606	19,259
Cooling		1,093	1,795	1,291									4,179
General appliance Load less heat pump	556	427	445	469	493	582	714	1,267	440	725	861	730	7,709
Total	1,000	1,520	2,240	1,760	880	1,240	4,429	4,827	3,391	3,680	3,844	2,336	31,147
Abingdon, Va.													
Heating	184	153			273	208	1,827	1,576	1,361	1,672	1,889	1,243	10,386
Cooling	143	545	1,208	847	226	91							3,060
General appliance Load less heat pump	913	913	952	953	941	621	1,020	832	806	886	974	902	10,803
Total	1,240	1,611	2,160	1,800	1,440	920	2,847	2,408	2,257	2,558	2,863	2,145	24,249
Roanoke, Va.													
Heating	135	26			126	301	1,364	2,084	1,768	2,276	1,759	811	10,650
Cooling	65	522	1,067	884	94	83							2,715
General appliance Load less heat pump	280	212	293	276	260	256	272	280	317	240	281	276	3,246
Total	480	760	1,360	1,160	480	640	1,636	2,364	2,085	2,516	2,043	1,087	16,611
Coshocton, Ohio													
Heating	444	3				5,880		1,995	4,185	3,731	4,366	2,292	22,905
Cooling	17	853	1,430	1,026									3,326
General appliance Load less heat pump	419	344	448	414		2,146		793	461	1,005	832	731	7,593
Total	880	1,200	1,878	1,440		8,035*		2,788†	4,646	4,736	5,198	3,023	33,824
South Bend, Ind.													
Heating	445	88			273	658	2,331	2,915	2,952	3,490	3,043	1,984	18,179
Cooling	24	378	1,143	1,002	33	32							2,612
General appliance Load less heat pump	604	694	537	698	694	590	546	1,154	872	860	830	528	8,607
Total	1,073	1,160	1,680	1,700	1,000	1,280	2,877	4,069	3,824	4,350	3,873	2,512	29,398

* September 1 to December 13.

† Beginning December 13.

Table VI. The General Appliances and Heat Pump Maximum Average 30-Minute Integrated Kilowatt Demand and the Monthly Load Factor for the Five Residences

	Abingdon, Va.			Roanoke, Va.			Coshocton, Ohio			South Bend, Ind.		
	General Appliances Only	Heat Pump Only	Combined Load	General Appliances Only	Heat Pump Only	Combined Load	General Appliances Only	Heat Pump Only	Combined Load	General Appliances Only	Heat Pump Only	Combined Load
May 1949	15.0	4.8	15.0	4.0	4.7	8.0	3.9	8.8	9.6	6.2	5.1	8.5
June	10.0	4.5	12.0	3.0	4.9	7.2	4.2	6.0	8.6	6.4	5.4	11.1
July	8.0	4.5	12.5	4.1	5.3	9.1	3.6	6.0	9.2	6.1	5.3	9.3
August	9.0	4.8	13.3	4.6	5.2	9.3	4.5	5.6	9.1	6.6	5.5	10.0
September	13.0	4.3	13.0	4.0	4.7	8.2				6.4	5.1	9.2
October	11.0	4.3	11.0	4.2	4.6	8.0				5.9	5.2	6.3
November	10.0	4.0	13.2	3.7	6.9	8.0				8.3	6.6	12.0
December	9.5	3.5	14.1	4.6	7.3	8.7	6.6*	5.2*	11.4*	8.9	6.5	12.7
January 1950	12.5	4.5	13.0	4.0	6.8	10.2	5.1	9.4	12.8	7.4	6.6	12.2
February	11.5	6.9	15.9	3.0	7.2	8.8	4.8	8.8	11.8	7.7	6.6	10.5
March	12.0	6.9	12.0	3.6	7.0	9.0	3.8	9.2	12.2	12.0	6.6	14.5
April	11.0	5.6	11.0	4.0	4.6	6.0	5.1	9.2	11.6	9.9	6.6	15.8
Monthly Load Factors, Per Cent												
May 1949	8.2		11.1	9.4		8.1	14.4		12.3	13.1		17.0
June	12.7		18.6	9.8		14.7	11.4		19.4	15.1		14.5
July	16.0		23.2	9.6		20.1	16.7		27.4	11.8		24.3
August	14.2		18.2	8.1		16.8	12.4		21.3	14.2		22.8
September	10.1		15.4	9.0		8.1				15.1		15.1
October	7.6		11.3	8.2		10.8				13.4		27.3
November	14.2		30.0	10.2		28.4				9.1		33.8
December	11.8		23.0	8.2		36.6	27.8*		56.6*	17.4		43.1
January 1950	9.6		23.3	10.6		27.5	12.1		48.8	15.9		42.1
February	11.5		23.9	11.9		42.5	31.2		59.6	16.6		61.7
March	10.9		32.1	10.6		30.5	29.4		57.3	9.3		35.9
April	11.4		27.1	9.6		25.2	19.9		36.2	7.4		21.8
Yearly Average Load Factors	11.5		21.4	9.6		22.4	19.5		37.7	13.2		29.9

* From December 13 to January 1.

periencing very close to the average outdoor temperature for the heating period. On each of these four figures the calculated 30-minute average kilowatt demand and load factor with and without the heat pump are given. While these data are not extensive enough to be conclusive they definitely and most clearly show that, for the cases considered, the daily residential load factor during the winter period is materially improved by the heat pump.

Discussion of Data and Tentative Conclusions

1. A thorough and complete analysis of all the data obtained during the test period from May 1949 to May 1950 has not been made as yet. The data in this paper, although representing only a small part of the information obtained from the field tests, will serve to give an over-all picture of the results. It is planned to make additional information available at a later date. Some of the test data included, even though they had not been analyzed fully and evaluated, are given so as to make the information available for those interested in the problem who may wish to study further the heat pump load.

A description of the systems, giving some of the operation and maintenance difficulties as well as possible improvements of the heat pump operating cycle has been covered in articles listed in the references.

2. Inconsistencies, some of which have not been accounted for or explained, as yet, may be present in the data included in this paper. The annual kilowatt-hours for heating was about 20 to 30 per cent higher than expected. In the case of Kingsport, this increase was partly due to the fact that storm windows were not installed until late in the heating season and to the unusually high losses existing in the duct system. In the case of Roanoke, part of the higher kilowatt-hours is undoubtedly due to the exposed uninsulated ducts in the basement and garage area. It is hoped that with further and more com-

Table VII. The Demand, Kilowatt-Hours, and Load Factor of the General Appliances and of the Combined Load for Four of the Installations During the Heating Cycle When Experiencing a Minimum, Average, and Maximum Outdoor Temperature

	Temperature, Degrees Fahrenheit		Demand		Kilowatt-Hours		Load Factor	
			General Appliances Only	Combined Load	General Appliances Only	Combined Load	General Appliances Only	Combined Load
Coshocton								
Friday	20 Min 52 Max	Average 36	3.5	9.5	21	130	25.0	52.0
Saturday	39 Min 51 Max	Average 45	4.0	10.0	28	154	29.2	64.2
Sunday	43 Min 57 Max	Average 50	1.1	9.3	2	92	7.6	41.2
Roanoke								
Saturday	31 Min 49 Max	Average 40	2.2	6.0	7	62	14.5	43.1
Thursday	36 Min 52 Max	Average 44	3.6	5.8	14	72	16.2	51.5
Saturday	31 Min 71 Max	Average 51	3.1	6.9	19	48	25.5	29
South Bend								
Thursday	13 Min 24 Max	Average 19	2.5	8.6	20	144	33.3	69.8
Saturday	26 Min 42 Max	Average 34	3.6	8.8	22	131	25.4	62.0
Sunday	40 Min 64 Max	Average 52	6.2	7.8	38	74	25.5	39.5
Abingdon								
Friday	31 Min 49 Max	Average 40	6.8	11.0	57	127	37.7	48.1
Saturday	29 Min 58 Max	Average 44	9.3	9.3	41	101	18.4	45.3
Thursday	38 Min 70 Max	Average 54	8.5	10.3	57	97	28	40.4

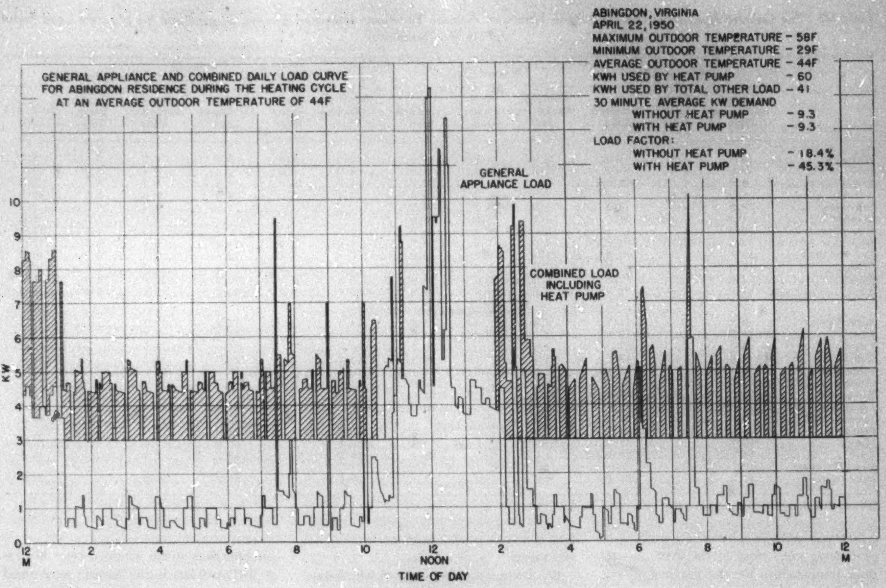


Figure 5. General appliance and combined daily load curve for Abingdon residence during the heating cycle at an average outdoor temperature of 44 degrees Fahrenheit

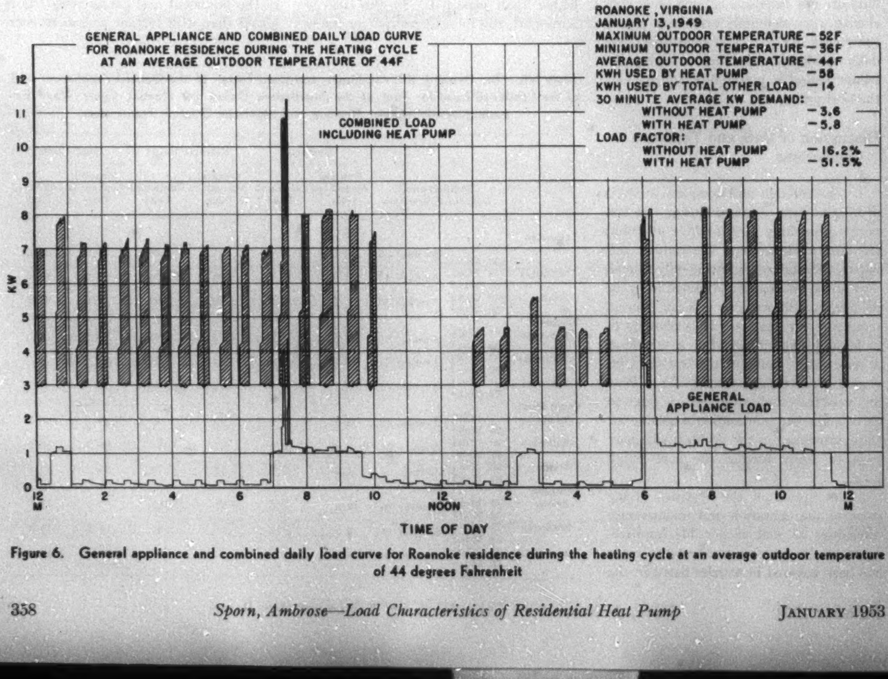


Figure 6. General appliance and combined daily load curve for Roanoke residence during the heating cycle at an average outdoor temperature of 44 degrees Fahrenheit

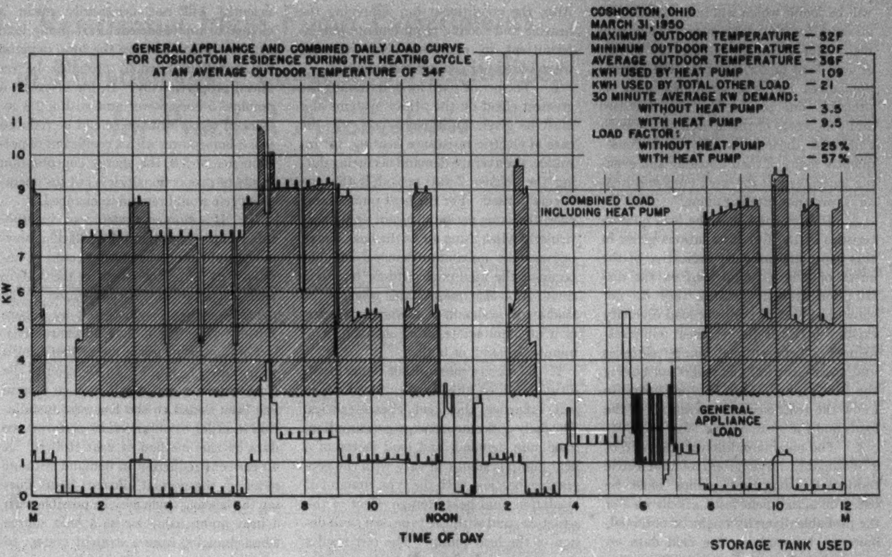


Figure 7. General appliance and combined daily load curve for Coshocton residence during the heating cycle at an average outdoor temperature of 36 degrees Fahrenheit

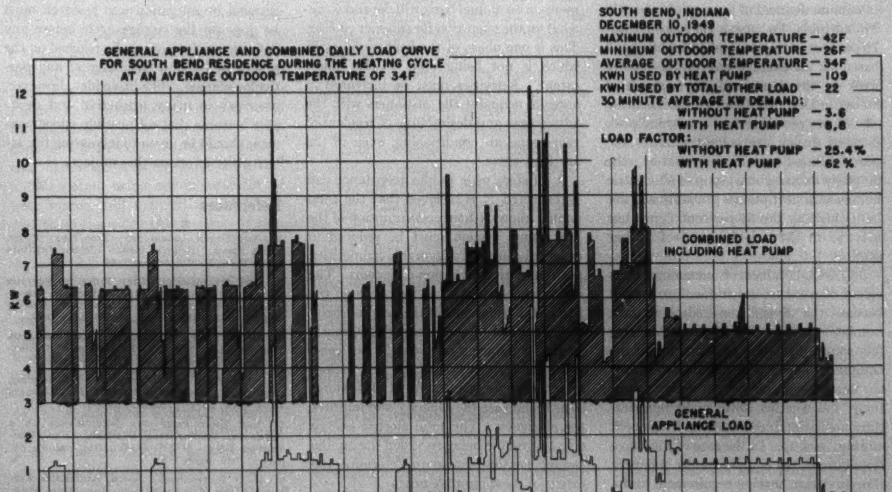


Figure 8. General appliance and combined daily load curve for South Bend residence during the heating cycle at an average outdoor temperature of 34 degrees Fahrenheit

plete analyses of the data, other factors will be found which will help explain and account for the differences. The main reason for Roanoke having the lowest annual load factor, for both the general appliances and the combined load, is probably due to the fact that the two adult members of the household are both employed during the day and consequently many of their noonday and evening meals, except for week ends and holidays, were not eaten at home.

3. It is realized that the load characteristics of the five installations given in this paper are not fully indicative of the effect of a heat pump load on the distribution systems because they do not show the effect of the peak load diversity existing between different heat pump residences or between heat pump residences and other electrical customers not having the heat pump. Also, it is necessary to know the relationship between all of the loads and the system peak.

4. The maximum kilowatt demand of a reasonably large group of all-electric residences using heat pumps must be based on actual field tests. Some idea of the probable diversity might be obtained, however, by studying the field data on conventional heating systems. The maximum demand of a group of houses in Oregon, using electric resistance heating, averages 54 per cent of the total individual demand. On the other hand, the maximum demand of a group of houses in Philadelphia, Pa., using gas heating, was 70 per cent of the total individual demand based on an average day of the coldest week in the year, and for a group of houses in Detroit, Mich., using oil heating was 67 per cent based on a morning peak demand during the coldest week of the year.[5] Based on this information, the diversity demand on a group of all-electric houses with heat pumps probably will not be as high as the 54-per-cent figure but may be in the neighborhood of 70 per cent.

5. Actual values of maximum coincident demand are very difficult to predict because the diversity will depend on a number of factors. Some influencing factors are the extent of oversizing of the equipment above that required to satisfy the heating requirements at design conditions, and whether lower night temperatures are maintained, thus requiring quick morning pickup. In the case of these five installations, the equipment size was very close to design heating requirement and a fairly uniform temperature was maintained throughout the 24-hour period. Also, the conditioner fan supplying the heating and cooling requirements was intermittent in operation—only running when heating or cooling was required.

6. The period used for determining the greatest effect on the utility systems also must be given additional study. In the case of electric resistance heating, for instance, the average demand occurring during the coldest 7-day period is the one normally used. For the heat pump, however, the same period may not apply particularly when using air as the heat source because the maximum demand will not occur at the minimum outdoor temperatures. It is also questionable whether the load analysis should be made on the basis of a 30-minute integrated demand, a 5-minute demand, or both.

7. There are many other items in addition to those already mentioned which will influence the load characteristics. The daily and yearly energy consumption, maximum demand, and load factor of a heat pump system will vary with the geographic location, with the type, amount of insulation, and heat storage effect of the structure, and with the type, size, and design of the heat pump. The test results show that even in the northern geographic areas the heat pump improves the load characteristics and surely this will hold true to a greater degree in the southern areas. The degree of load factor improvement on an annual basis will depend somewhat on the operation for summer cooling. This is one phase of the heat pump cycle which is not being used to the fullest extent. Activities must be increased in order to acquaint the customers with the advantages and desirability of residential year-round air conditioning, even in the northern areas.

8. Before wide public acceptance can be expected, it is believed that the total annual kilowatt-hours consumption of the heat pump system must be reduced in order to be in a more favorable competitive position with conventional systems. The average coefficient of performance on the five installations was in the neighborhood of 2.3 instead of the expected 3.0, representing a reducing of about 20 per cent. After the termination of the test period, design changes were made in all systems so as to increase the capacity and improve the efficiency. The total electric energy consumption is being recorded during the coming heating season to determine the extent of these improvements.

9. The reduction of the total kilowatt-hours due to increase in equipment efficiently will not necessarily mean a change in the residential heat pump load characteristics, because the total demand per installation would probably be reduced proportionally. If, for instance, a nominal 5-horsepower unit with a 2.3 coefficient of performance could be reduced to a 3-horsepower with a coefficient of performance of 3.8, the energy consumption would be proportionally less but the operating cycle would remain unchanged.

10. If a storage system was incorporated in the heat pump installation, however, the load characteristics could be improved as far as the effect on the system peak is concerned. An indication of the possible change is illustrated by Figure 7 which gives the daily load characteristics of the Coshocton system when using a water storage system. In addition to the Coshocton system, a salt storage system has been added to the Roanoke installation in order to obtain some performance data on this method of heat storage. A 40-per-cent reduction in demand without reducing the normal kilowatt-hours during the heating cycle may be possible with a heat pump using air as a heat source when changing from a straight system to one using storage. It is also entirely feasible to have an almost flat load curve by selecting a proper tank size and having suitable controls to limit the current consumption during the peak period. Additional investigation and research must be done on the storage cycle before any definite conclusions can be reached on the practicability and feasibility of any particular scheme. The basic idea, however, possesses so many attractive and desirable features that all possible encouragement should be given to increasing the activity and effort on this subject.

References

1. The Sixth Ingredient—The Heat Pump, **Philip Sporn.** *Bulletin*, Edison Electric Institute (New York, N. Y.), August 1944.

2. The Residential Heat Pump in Various Climates for Summer and Winter Air Conditioning, **Philip Sporn, E. R. Ambrose.** *Heating and Ventilating* (New York, N. Y.), May 1949.

3. Heat Pump Undergoing Tests in Five Houses, **Philip Sporn, E. R. Ambrose.** *Architectural Record* (New York, N. Y.), May 1949.

4. Heat Pump Installation for Residential Use, **E. R. Ambrose.** *Mechanical Engineering* (New York, N. Y.), February 1951.

5. The Heat Pump—Its Significance as a Potential Residential Electric Load, **Constantine Bary.** *AIEE Transactions*, volume 68, part I, 1949, pages 1–8.

6. Heat Pumps (book), **Sporn, Baumeister, Ambrose.** John Wiley and Sons, Inc., New York, N. Y., page 155.

No Discussion

Residential Heat Pump Experiments in Philadelphia—Suggested Possibilities for Practical Applications

CONSTANTINE BARY
MEMBER AIEE

Synopsis: As a result of operating experience on experimental heat pumps in Philadelphia, this paper suggests the necessity for improving the annual load factor of heat pump systems through the employment of heat storage. It demonstrates the relationship between minimum capacity of heat storage with reduced capacity of heating plants, and concludes that it should be economically practical to supply electricity to heat storage types of heat pump systems in competition with other fuels.

IT IS generally conceded that at the existing state of the art and under conventional methods of operation electricity cannot compete in cost of operation for house heating on a direct energy conversion basis with other fuels such as coal, oil, or gas, customarily used.[1] The basic reasons for this handicap lie in the relatively large factor of kilowatt-hour use per 1,000 cubic feet per degree day for any resistance type of house heating, and the inherent low annual load factor of house-heating requirements in most regions of the United States.[2] The heat pump in theory can and, as has been demonstrated by actual field installations, does overcome to a large extent the first feature of that handicap. Practice has shown that a heat pump, using air or earth for its heat source, will multiply between two and three times the effectiveness of electricity for heating purposes. It thus produces for a given job of house heating the desired effect of reducing substantially the quantity of electricity required by any now known form of resistance type of heating. However, the magnitudes of the over-all performance factors attained in practice on a number of residential heat pump installations,[3] or that are indicated to be feasible from economic considerations, are insufficient to overcome completely the competitive handicaps previously mentioned, especially in view of the relatively high initial cost inherent in residential heat pump installations.

It has also become clear, from actual experience that:

1. In climates similar to Philadelphia, the application of heat pumps for home heating and cooling under conventional methods of operation cannot be expected to improve appreciably the annual load factors of the electrical load of this device over the inherently low annual load factor of heating requirements.[2]

2. The electric load of the heat pump under conventional methods of operation does not offer possibilities for any substantial improvement in the annual load factor of the residential class of customers.[1]

Since, with other things equal, cost per kilowatt-hour of supplying electrical service varies inversely with load factor in accordance with the laws of an equilateral hyperbola, and in view of the foregoing statements, it would be impractical to expect a lower average rate per kilowatt-hour to the heat pump load than is obtained now from the addition of uncontrolled electric loads of other major uses in the home. As a matter of fact, under certain conditions, to cover all costs for the load factors of the residential class obtained with heat pumps operating under conventional methods, the average rate per kilowatt-hour applicable to this device may have to be even higher than is now generally available at prices of the end blocks of residential rates. These prices now average close to 2-cents per kilowatt-hour in the 16 largest cities of the United States, located in the Northern and Middle Temperature zones.

As a result of actual experience in the operation of experimental heat pump systems in Philadelphia[3] and the previously outlined operating and economic considerations, it has become evident to the author that it is necessary to direct efforts toward finding means for improving substantially the annual load factor of the electric service supplied to a heat pump system to overcome the competitive handicaps previously mentioned. It is thought that this objective may be accomplished through the development of economically feasible media for storing heat in sufficient quantities to permit a substantial reduction in the maximum demand for electricity of a heat pump system. Obviously this would introduce new and unconventional methods of operations in comparison with existing practices, but if successful it should go a long way toward bringing the heat pump (and thus the use of electricity for home heating) closer to a commercial reality demanded by the public and sought by electrical utilities and manufacturers of equipment, in conformity with the basic principle of the American system of enterprise that the luxury of today, used by a few, becomes the necessity of tomorrow, required by many.

Object

From an analysis of actual weather cycles, the theoretical quantitiative relationships which determine the minimum capacity of a heat storage system required with capacity of the heating plant reduced below the required value under conventional design conditions will be demonstrated. Some basic methods of supplying heat to a heat storage system in conjunction with heat pumps will be illustrated, and the effects of these applications upon the load characteristics of the electrical service supplied to heat pump systems will be shown.

House-Heating Requirements

Climatological data employed in this paper are those recorded by the United States Weather Bureau for Philadelphia. Thus, results and conclusions contained herein, unless otherwise qualified, are confined to conditions there and to other territories with similar climates.

Figure 1 depicts the theoretical daily Btu heating requirements of a dwelling on the basis of hourly changes of outdoor temperatures and a fixed indoor temperature of 70 degrees Fahrenheit. They are confined to the period from the beginning of December to the end of February because it was found, from practical considerations, that the months before December and after February impose heating requirements, even for the coldest season on record, that can be satisfied by the lowest capacity of the heating plant and the corresponding size of storage plant used in this paper. For

---

Paper 52-340, recommended by the AIEE Domestic and Commercial Applications Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted January 16, 1952; made available for printing September 3, 1952.

CONSTANTINE BARY is with the Philadelphia Electric Company, Philadelphia, Pa.

The author acknowledges the valuable assistance and suggestions of A. Abramovitz of the Philadelphia Electric Company in the preparation of the material contained in this paper.

convenience of application, all data in this figure are expressed per 1,000 Btu per hour of design heat requirements at zero degrees Fahrenheit outdoor temperature.

The plotted data in this figure were developed from an analysis of the hourly temperature changes during the 1917–1918 heating season, the coldest recorded in Philadelphia, see Figure 1(A), and during the 1934–1935 heating season which can be taken as representative of a "normal" or average heating season there, see Figure 1(B). For convenience of calculations and for purposes of this demonstration, the relationships between degree-hour deficiency D and hourly heating requirements R were calculated from the following equation for a heat loss of 1,000 Btu per hour for the indicated standard design conditions of indoor and minimum mean daily outdoor temperatures:

$$R=\frac{1{,}000 \text{ Btu/hour}\times D}{65 \text{ degree-hours}}$$

These calculations were carried out for each hour of the selected period. The sum of the hourly requirements for each day of the period are depicted in these curves, and the shaded areas on them depict the net deficiency for the day in the ability of the indicated size of the heating plant to meet the stated heating requirements. For purposes of this illustration, the capacity of the heating plant has been at two-thirds of the maximum hourly design heat requirements.

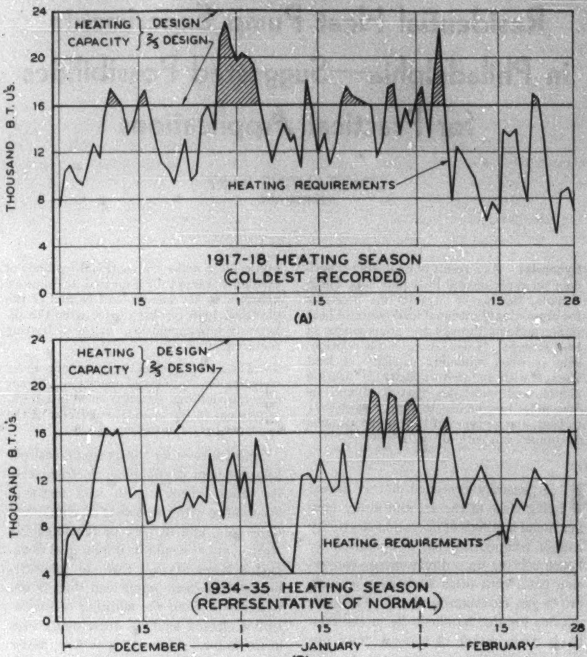


Figure 1. Illustrative daily heating requirements for a 1,000-Btu-per-hour design heat loss based on United States Weather Bureau climatological data for Philadelphia

Heat Storage Requirements

Figure 2 shows two sets of derived theoretical relationships between a given value of capacity of a heating plant and the minimum capacity of heat storage installation required, and between capacity of heating plant and the total quantity of heat that must be supplied by the storage system so that the heating plant and the storage system together satisfy the entire heating requirements of the house at all times during a stated heating season. Curves *A-1* and *B-1* depict these relationships for the 1917–1918 heating season, and curves *A-2* and *B-2* for the 1934–1935 heating season. Both sets of curves were developed from the hourly heat requirement data which form the basis of the curves depicted in Figure 1. It should be noted that these relationships are purposely confined to the theoretical values and, therefore, do not include the additional heat requirements due to stand-by or radiation losses of the storage system, nor due to other practical conditions which would be present in actual installations and which may result in heat storage requirements that may be more or less than indicated by the theoretical relations.

It will be observed from a study of Figure 2(A) that for capacities of heating plants in excess of 65 per cent of the design heat requirements, the theoretical capacity of the storage installation, which is necessary to provide enough stored heat at the beginning of December to meet the requirements of the coldest spells of the heating season (exclusive of stand-by losses and so forth) represents a relatively small portion of the total heat requirements of the premises for a normal heating season. At a 65-per-cent capacity value of the heating plant, it is about 2.14 per cent of such requirements for the coldest recorded heating season (curve *A-1*) and about 1.2 per cent for a normal heating season (curve *A-2*). A decrease in such capacity, however, substantially below the 65-per-cent value is accompanied by an accelerating increase in the theoretical capacity requirements of the heat storage plant.

The relationships between the total quantity of heat that has to be supplied through the storage system during a heating season and the stated capacity of the heating plant (Figure 2(B)) generally follow the minimum capacity requirements of the storage installation. But, as may be expected, they are at a somewhat higher numerical value. At 65-per-cent capacity of the heating plant at design conditions, the total quantity of heat that has to be supplied from storage, in relation to the total heat requirements of the premises for a normal heating season, is about 4.4 per cent in the coldest season on record and 1.9 per cent in a normal one.

From considerations of heat pump application to house heating, a reduction in its capacity for heating below the conventional values of design requirements is limited by the capacity it has to have to satisfy the conventional design requirements for cooling. For climatic conditions similar to those encountered in Philadelphia, the relationship between these capacity requirements is in the order of three for heating to two for cooling. Thus, insofar as that climate is concerned, the maximum possible improvement in the annual load factor of electrical service supplied to a heat pump system is limited to the provision of a heat storage system of a size that would be sufficient to meet the total heating requirements of the house with a heat pump capacity large enough to meet the conventional design *cooling* requirements of the premises. This would require a heat pump about 65 per cent of the size which would be provided under conventional design heating requirements.

To minimize the physical size of the heat storage installation, research is now in progress[4,6] to obtain an economical, nontoxic, noncorrosive and noninflammable material (with a high specific heat and a high unit value of heat of fusion) which would alternately store and, when required, release the stored heat under specified temperature conditions in a smaller space than is possible with water as a storage medium. The problem is largely of a physicochemical nature. It is mentioned here to call to the attention of industries operating in that field and to others the importance of the development of such a material.

Depending on the material developed and the temperature ranges involved, there will be a variety of methods which may be used in the supply of energy to actual heat storage installations. In this paper mention will be made of two basic methods which can be used in conjunction with a heat pump installation.

Table I. Illustrative Significant Features of Calculated Load Characteristics of Electrical Service Supply to an Average Heat Pump System

			Heat Storage Type With Heat to Storage Supply by	
	Conventional Type	Ideal Type	The Heat Pump (Method Number 1)	Resistance Elements (Method Number 2)
Calculated heat requirement—Btu per hour standard conditions 0 degree Fahrenheit outdoors, 70 degrees Fahrenheit indoors	55,000	55,000	55,000	55,000
Maximum heat output of heat pump installation at 0 degrees Fahrenheit—Btu per hour	85,000	55,000	35,750	35,750
Electric load characteristics of heat pump system				
Maximum 30-minute demands—kilowatts				
Noncoincident for year's cold spell	8.2	5.4	3.5	3.5
Diversified on day of 0-degree-Fahrenheit mean temperature	6.8	5.4	3.5	3.5
Diversified on average day of normal coldest week	5.7	4.0	3.5	3.5
Annual energy for normal year—kilowatt-hours				
For heating requirements				
By heat pump directly	9,550	9,550	8,850	8,600
By storage system			750	2,900
Total for heating	9,550	9,550	9,600	11,500
For cooling	1,000	1,000	1,000	1,000
Total for heating and cooling	10,550	10,550	10,600	12,500
Coincidence factors—per cent				
Based on diversified maximum demand on day of 0-degree-Fahrenheit mean temperature	83	100	100	100
Based on diversified maximum demand on average day of normal coldest week	70	74	100	100
Annual load factors—per cent				
Based on noncoincident maximum demand	14.7	22.4	34.6	40.8
Based on diversified maximum demand				
On day of 0-degree-Fahrenheit mean temperature	17.7	22.4	34.6	40.8
On average day of normal coldest week	21.0	30.2	34.6	40.8

Supply of Heat to Storage System

The first method, illustrated in a schematic diagram in Figure 3(A), employs the heat pump itself for the supply of heat to the storage system when needed and at times when the heat pump is not called upon to furnish the heat requirements of the house.[5] Obviously, this method is confined to the temperature levels economically feasible with heat pumps, but it possesses the advantage of providing the heat for storage at the multiplying factor which the heat pump is capable of producing to increase the effectiveness of electricity for heating purposes. It has the disadvantage of mechanical complications of additional heat transfer surfaces required.

The second method, on which a patent has been applied for, is illustrated in a schematic diagram in Figure 3(B), and employs electric resistance elements to supply heat to the storage system when needed and at times when the heat pump is not called upon to furnish the heat requirements of the house. The total connected load of resistance elements is such as not to exceed the maximum demand of the heat pump during the coldest portion of the heating season, and these elements are so interlocked electrically with the heat pump that they cannot be in operation at the same time with the heat pump. This method need not be confined to the temperature levels economically feasible with the heat pump, but the heat supplied to the storage system is obtained from electricity under a direct energy conversion basis. This latter feature, which is accompanied by a lower rate of heat recovery than is possible with a heat pump, introduces larger capacity needs for the

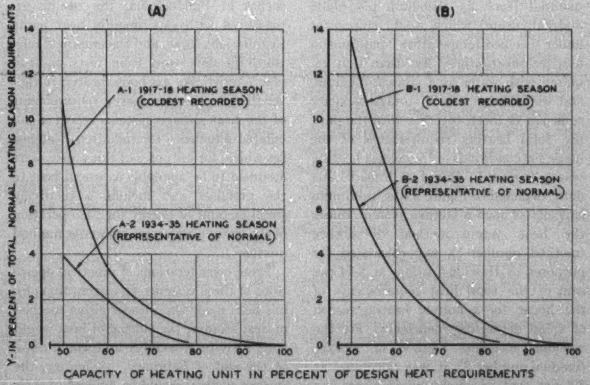


Figure 2. Relation of capacity of heating unit to (A) minimum capacity of heat storage installation required and to (B) total quantity of heat required from the storage system during the heating season based on U. S. Weather Bureau climatological data for Philadelphia

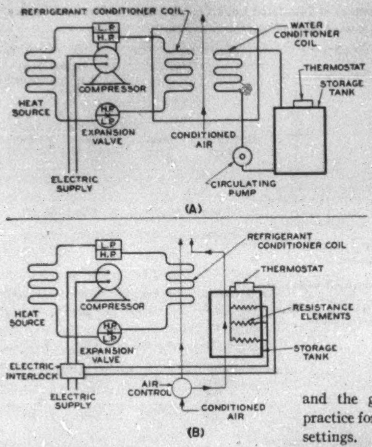


Figure 3 (left). Schematic diagram of heat pump system using storage

(A) Heat to storage by the heat pump

(B) Heat to storage by resistance elements

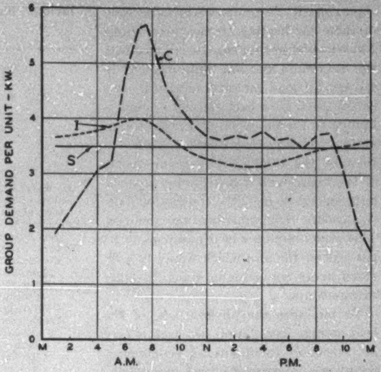


Figure 4 (above). Illustrative load curves for the average day of the normal coldest week for an average residential heat pump system (based on 1/2-hour integrated demands)

C = Conventional type (reflecting night cutback in thermostat setting)

I = Ideal type (fixed thermostat setting day and night)

S = Heat storage type with capacity of heat pump unit at 65 per cent of design requirements

storage system than is required (with other conditions equal) with the heat pump method of heat supply. However, as a result of higher temperature levels possible with this method, the physical size of the storage installation need not be any larger than that provided under the first method. The resistance elements method does not possess the disadvantages of the mechanical complications of the heat pump method, and has the definite advantage of providing a higher degree of reliability in the supply of heat to the house, since it incorporates a second source of heat production which, in an emergency, could be used to meet some or all of the heating requirements of the house in case of a failure of the heat pump unit.

Load Characteristics of Electrical Service

Table I illustrates the significant features of calculated load characteristics of electrical service supply to an average heat pump system for the indicated design heat requirements and for a normal heating and cooling season in Philadelphia.

Electric energy and demand requirements of heat pump units in the table reflect a performance factor of three, which is believed feasible of attainment, on the basis of recent experience on experimental heat pump installations in Philadelphia.[5]

The designation "conventional type" refers to installations which would approximate conditions generally encountered in present-day gas or oil house-heating installations, embracing the provision of generous margins in capacity of units and the generally prevailing operating practice for a night cutback in thermostat settings. The designation "ideal type" refers to installations which would reflect the theoretically ideal, but practically attainable, conditions comprising the feature of equality between heat pump capacity and calculated design heat requirements, and the practice of operating under a fixed thermostat setting day and night during the heating season. A more detailed description of these types of installations is contained in the author's paper "The Heat Pump Load—Good or Bad?"[6]

The designation "heat storage type" refers to installations which would employ a heat pump unit with a heating capacity of 65 per cent of calculated heat requirements under standard conditions, and a heat storage system containing a material (such as, disodium phosphate dodekahydrate) which would store heat under specified temperature conditions in a space substantially less than that required with water as a storage medium, and would have a heat storage capacity sufficient to meet, together with the unit, the total heating requirements of the house for conditions of the coldest heating season on record. Based on Philadelphia data, depicted in Figure 2, the minimum capacity of such a storage system, under the heat pump method of storage (method number 1), was calculated, for purposes of this illustration, at 2.14 per cent of the total heat requirements of the house for a normal heating season of 4,750 degree-day deficiency. For the resistance-elements method of storage (method number 2), it was calculated at 3.3 per cent of such requirements, to reflect a calculated allowance for the substantially lower rate of heat recovery possible under this method in contrast with the heat pump method of heat supply to storage.

Data on electrical demands and energy characteristics in the table were derived from the information contained in the author's paper previously mentioned;[6] they reflect the assumed conditions used in this paper for house-heating requirements in Philadelphia, the results experienced with experimental heat pump installations there, and the empirical relationships developed from tests on load characteristics of gas house heaters in that territory.[7] Electric energy requirements of the heat storage systems reflect a calculated allowance for radiation heat losses which, for purposes of this paper, were assumed to be available as useful heat to the premises and include a relatively small amount of energy for the operation of the water pump required under method number 1.

From considerations of practical operations of these systems, an assumption also has been made that, under either method of heat storage, the storing of heat would start early in November and would cease at the end of February, thereafter, the storage systems would be drained of their stored heat by employing it for supply to the premises before the expiration of the heating season.

Data in Table I show that for a normal heating season, and depending on the method used for storing heat, a storage system requires between 7 and 23 per cent of the total annual electric energy taken by the heat pump installation. They also show that, although the total amount of electric energy required under the resistance element method of heat storage is about 18 per cent higher than under the heat pump method, the annual load factors under the former method also are about 18 per cent higher than under the latter one. With a properly designed demand-energy form of rate, however, the added cost of electric energy with resistance-type storage over that with the heat pump would be relatively small, in the order of less than 5 per cent of the total annual bill for the energy used by the heat pump system. Of utmost significance, too, is the fact that under either of these methods, the heat storage type of installation produces annual load factors of electrical service supply which are substantially higher than is possible with either the conventional type or even with the ideal type reaching levels which either approach or exceed those produced by the uncontrolled load of the residential class as now constituted.[1]

Figure 4 depicts illustrative load curves for an average day of the normally coldest week in Philadelphia of electrical service supplied to an average heat pump system for each of the three types described in this paper. It is of interest to note that the heat storage type of installation will produce, on such a day, a continuous and uniform electric load over the entire 24-hour period. Wherever that load curve falls below the load curve of the ideal type, indications are that the deficiency in heat supply to the house by the heat pump is made up from storage; and wherever it lies above the load curve of that type, the difference between these curves represents the amount of electricity used for replenishing the heat withdrawn from the storage system.

Major Features of Heat-Storage-Type Heat Pump Systems

From the illustrative data on electric load characteristics contained in Table I and Figure 4 it should be noted that, as a result of the higher annual load factors of the heat storage type of heat pump installation, the equipment of that system will be subject to longer hours of operation in a year than under either the conventional or ideal types. Thus the design and production of such units must embrace this feature, by providing equipment that can withstand these long hours of operation over many years of its use, in order to satisfy the accepted reliability requirements of other conventional heating and cooling systems to which the public is accustomed. Also it should be noted that this type of heat pump system will operate almost continuously for the major part of the winter season. This probably will result in some advantages from mechanical considerations of the lubrication and refrigerant cycles, as well as of the comfort conditions of continuous air circulation in the home. But in those cases where the earth is considered for the heat source for such heat pump installations, the continuous nature of operation of this type of system over a long period of time during the coldest period of the winter months introduces the necessity for more and better information on heat extraction features from the earth with accompanying freezing under substantially steady conditions of heat withdrawal from the earth over such long periods of time. Test results obtained on this subject from the experimental heat pump installations in Philadelphia, for example, cannot be considered fully conclusive on this matter because of the relative shortness of the freezing period. But, for the conditions actually encountered there which reflected continuous and long hours of heat pump operation during the winter cold spells, the results are encouraging for the heat pump using earth for its heat source.[8]

With the reduction in the heating capacity of the heat pump possible through the use of heat storage, air becomes a more attractive source of heat for the heat pump in such climates as Philadelphia. But the basic handicaps of that source are not eliminated through the use of heat storage described in this paper; they are simply decreased, perhaps, to a lower level of capacity considerations.

Conclusions

From the foregoing discussion it is possible to draw the following six conclusions:

1. The employment of heat storage in conjunction with a heat pump offers the opportunity of reducing the size of the unit from its conventional design requirements.

2. A heat storage type of heat pump installation increases the annual load factor of the electrical service supplied to the system.

3. The increase in annual load factor of electrical service supplied to a heat storage type of heat pump system is dependent inversely upon the decrease in the size of the heat pump unit below design heat requirements. A substantial reduction in such a size is required for a substantial increase in the annual load factor of the heat pump system. In practice such decrease is limited by either the capacity requirements of summer cooling or by the economical size of the heat storage system.

4. A development of materials, which would permit an economic and reliable storage of heat, is essential for the commercial acceptance of heat pump application for house heating in climates such as encountered in Philadelphia.

5. As a result of the desirable levels of annual load factors feasible with heat storage type of heat pump systems and under performance factors possible of realization, it should be economically practical to supply electrical service to such systems at a price that would enable electricity to compete in cost of operation for house heating with other fuels such as coal, oil, or gas, customarily used.

6. Electrical service to heat pump systems should be supplied under a demand-energy form of rate, which is generally accepted as being able to reflect the benefits of high load factor operation.

References

1. The Heat Pump—Its Significance as a Potential Residential Electric Load, **Constantine Bary.** *AIEE Transactions*, volume 68, part I, 1949, pages 1–8.

2. Electric House Heating Load Characteristics and Economics, **W. F. Friend.** *Proceedings*, Midwest Power Conference (Chicago, Ill), April 1951.

3. Residential Heat Pump Experiments in Philadelphia—Installation and Operating Experience, **J. H. Harlow, G. E. Klauper.** *AIEE Transactions*, volume 71, part II, 1952, pages 366–75.

4. Storing Solar Heat in Chemicals, a Report on the Dover House, **M. Telkes.** *Heating and Ventilating* (New York, N. Y.), November 1949.

5. A Report on Heat Storage for Heat Pumps, **Joint AEIC-EEI Heat Pump Committee.** *Bulletin*, Edison Electric Institute (New York, N. Y.), September 1950.

6. The Heat Pump Load—Good or Bad? **Constantine Bary.** *Refrigerating Engineering*, American Society of Refrigerating Engineers (New York, N. Y.), May 1950.

7. Characteristics of Gas House Heating Load, **Constantine Bary.** *American Gas Association Monthly*, American Gas Association (New York, N. Y.), March 1949.

8. Residential Heat Pump Experiments in Philadelphia—Earth as a Heat Source, **A. H. Kidder, J. H. Neher.** *AIEE Transactions*, volume 71, part II, 1952, pages 343–50.

No Discussion

Residential Heat Pump Experiments in Philadelphia—Installation and Operating Experience

J. H. HARLOW
MEMBER AIEE

G. E. KLAPPER
NONMEMBER AIEE

Synopsis: In order to investigate the practicability of using heat pumps for year-round residential heating and cooling in the Philadelphia area, two experimental installations were operated for 2 years, using earth as a heat source. This paper describes the installations, presents essential performance data, and concludes that the satisfactory performance experienced may be bettered by specific technical improvements.

SINCE 1946 extensive interest has been shown throughout the electrical utility industry, and by the public, in the possibilities of a practical method to heat the home electrically, with the accompanying advantages of cleanliness, absence of combustion, ease of operation, and general comfort and convenience. It has become increasingly apparent that, in the present stage of the art, the only practical method now available for using electricity for house heating is to use a device which would multiply the effectiveness of electricity for heating purposes by several times its direct energy conversion value. The heat pump is such a device that accomplishes the desired effect through well-known thermodynamic principles. To assist the commercial development of this device, and to obtain first-hand knowledge of its engineering, economic, and operating features, in 1947 the Philadelphia Electric Company undertook to establish a joint co-operative program with a manufacturer of air-conditioning equipment for the installation, in its service area, of two experimental heat pump units to obtain as much as possible of factual information on, and experience in, the operation of this device.

Installation

SELECTION OF EQUIPMENT

Investigation of the products of several manufacturers then producing heat pumps indicated that the heat pump assemblies they proposed to furnish were not satisfactory to meet the requirements of the experiment. Discussions with other manufacturers of refrigerating equipment indicated that they did not desire to enter into development of the heat pump at that time.

Finally an arrangement was made with Chrysler Air Temp Sales Corporation, whereby that company agreed to supply much of the mechanical equipment and to investigate equipment design and changes as indicated by operating experience. The Philadelphia Electric Company agreed to furnish all testing equipment and other parts of the installation, and to conduct all necessary performance tests on the units for a period of 2 years.

BASIC DESIGN PRINCIPLES

The adopted refrigeration system was of the compression type, using Freon-12 as the working fluid. The size of the compressor units made available by Chrysler was 5-horsepower; they were hermetically sealed, motor-compressor combinations. Each unit was equipped with automatic change-over valves for switching from the heating cycle to the cooling cycle. The earth was selected as the heat source with proper coils imbedded in the ground surrounding the houses, as it was felt that in the service area of the company it presented the most reliable and generally available source with relatively constant year-round temperatures. A mixture of ethylene glycol and water (hereinafter referred to as glycol solution) was selected as the liquid to be circulated through the ground coils, and a heat exchanger was used for heat transfer to the refrigerant. In each installation the glycol solution was circulated with a small electrically driven pump. This indirect method of heat transfer was selected in preference to directly expanding the refrigerant in the earth coils, to avoid the use of large quantities of refrigerant, to avoid the trapping of compressor lubricating oil in these pipes, and to provide a reliable and workable basis for determining the heat flow to and from the earth.

SELECTION AND DETAILS OF RESIDENCES

The two residences used in the experimental work were selected upon the following basis:

1. Owners to be employees of the Philadelphia Electric Company, with engineering training and an understanding of heat pump principles, to simplify operating procedure.
2. Homes to be heated with hot-air systems to permit easy adaptation of the heat pump installation.
3. Property surrounding the houses to be adequate in size for installation of the desired lengths of ground coils.

In this paper the two installations will be referred to as "Whitemarsh" and "Lansdowne."

The Whitemarsh installation comprised a 3-story stone-and-frame structure with partial basement, constructed in 1938, Figure 1. Exterior walls are of normal stone, insulated, and furred-plaster construction. Storm windows and doors were available and were used. All existing air supply and return ducts and baseboard supply and return grilles were used without alteration. The calculated heating requirement for this home was approximately 90,000 Btu per hour, based on standard conditions of zero degrees Fahrenheit outdoors and 70 degrees Fahrenheit indoors. An efficiency test run on the original gas-fired furnace (Bryant model *200A C78*, rated at 160,000 Btu per hour output) just before the heat pump installation, corroborated the calculated heating requirements. The cooling requirements for this home were calculated at 40,000 Btu per hour.

The Lansdowne installation comprised a 3-story frame structure with complete basement, constructed about 1928, Figure 2. Existing air supply grilles were located above baseboards, and central return grilles were located in the living room floor. Exterior walls were of normal wood, sheathed, insulated, and of furred-plaster construction. Storm windows and doors were available and were used. At the time of the installation of the heat pump, the round air supply ducts in the basement were replaced with rectangular ducts, and the original lattice-type air supply registers were replaced with modern adjustable vane-type registers. The central return air registers were used and return ducts were installed in the basement between these registers and the

---

Paper 52-342, recommended by the AIEE Domestic and Commercial Applications Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted January 16, 1952; made available for printing September 3, 1952.

J. H. HARLOW and G. E. KLAPPER are with the Philadelphia Electric Company, Philadelphia, Pa.

heat pump unit. The calculated heating requirement at standard conditions for this home was approximately 75,000 Btu per hour. An efficiency test, run on the original oil-fired furnace just before installation of the heat pump, corroborated the calculated heating requirements. The original heating system was a gravity-flow-type Holland hot-air furnace with an Esso conversion oil burner rated at $1^1/_2$ gallons per hour. The cooling requirements for this home were calculated at 35,000 Btu per hour.

DESCRIPTION OF HEAT PUMP UNITS

The heat pump units, supplied by the Airtemp Division of the Chrysler Corporation, were housed in the standard casing for its 5-horsepower self-contained air-conditioning unit. Separate fan and filter units were provided to circulate the air through the houses. The refrigerant system was completely housed in the casing, Figure 3. The schematic refrigerant cycle is shown in Figure 4.

The standard 5-horsepower hermetically sealed compressor, item 1, Figure 4, was modified to include an automatic unloading device which was capable of cutting any number of the five cylinders out of service to modulate the capacity of the unit with the heating requirements. The compressor pumped high-pressure, high-temperature Freon-12 refrigerant through the oil separator (3) to the automatic 4-way valve (4). This valve was automatically positioned by the room thermostat to satisfy either a heating or cooling requirement. In the heating position, the refrigerant then was supplied to the combination heating-cooling coil (12). This special coil had a header so arranged that the coil could be used as either an evaporator or a capacitor. When used as a capacitor, air from the house was blown over this coil to pick up the heat of condensation of the refrigerant, and the hot air was supplied to the various rooms through ducts. The refrigerant having condensed to a high temperature, high pressure liquid then flowed to the receiver (9). Then it flowed through a liquid subcooler (8), a strainer ([illegible]), and into the high-pressure float valve (6). This valve fed the cool liquid refrigerant into the glycol heat exchanger (5). Being on the low-pressure side of the float valve, the refrigerant evaporated in this heat exchanger, getting the heat of evaporation from the glycol solution which was being circulated through pipes buried in the earth. This solution then circulated back to the earth coils at a temperature below the earth ambient temperature, thus picking up more heat from the earth to be extracted in the heat exchanger. The cool refrigerant vapor then circulated through a coil in the liquid subcooler (8). This subcooler served a twofold purpose: first, to cool the liquid refrigerant prior to expansion, thus improving the refrigerating cycle efficiency, and second, to superheat slightly the refrigerant returning to the compressor, thus minimizing the possibility of liquid refrigerant entering the compressor and causing damage. From the subcooler the gaseous refrigerant returned to the suction of the compressor to repeat the cycle.

Thus heat was transferred from the earth to the glycol solution circulating in the buried pipes. This heat then was transferred to the refrigerant, and its temperature was raised to a point where it could be usefully extracted for human comfort (110 to 140 degrees Fahrenheit). The heat then was transferred at the heating coil from the refrigerant to the air circulating to the house.

During cooling cycle operation the 4-way valve reversed the flow of the refrigerant, and heat was transferred from the house to the earth.

In the original design the oil cooler, item 2, Figure 4, was so connected that some of the cool glycol solution from the earth was by-passed through the oil cooler to absorb this heat, which was produced mainly by compressor friction. While this heat was later recovered in the cycle, it was done at the expense of using part of the compressor capacity. Before the test got far under way, a new oil cooler was located in the return air stream. Thus, this heat was transmitted directly to the

Figure 1 (left). Whitemarsh

Figure 2 (left). Lansdowne

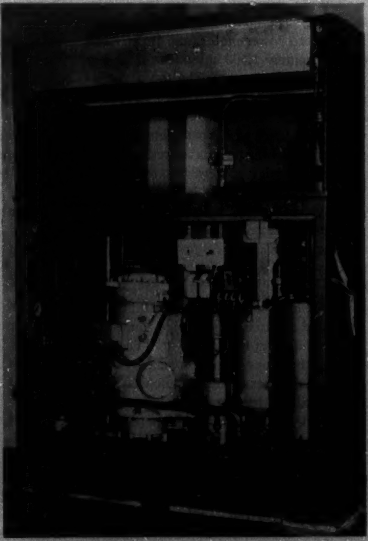
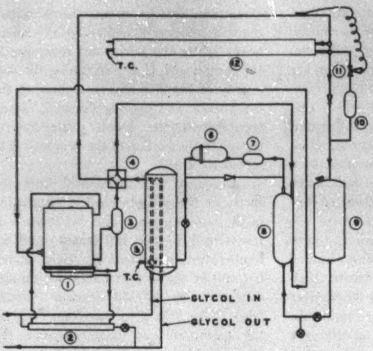


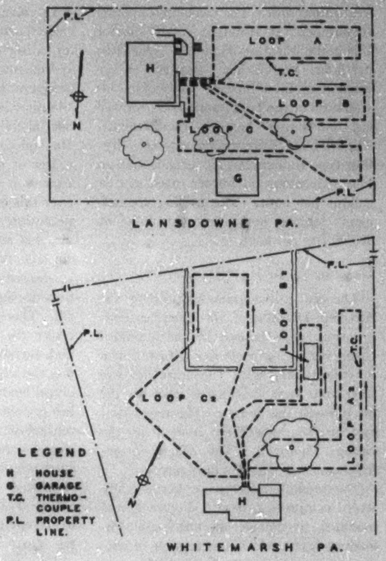
Figure 3 (right). Heat pump unit



ITEM	NAME	ITEM	NAME
1	5 H.P. COMPRESSOR	8	LIQUID SUBCOOLER
2	OIL COOLER	9	LIQUID RECEIVER
3	OIL SEPERATOR	10	LIQUID STRAINER
4	ALCO 4 WAY VALVE	11	EXPANSION VALVE
5	GLYCOL COOLER	12	KENNARD COIL
6	H.P. FLOAT VALVE	T.C.	THERMOCOUPLE.
7	STRAINER		

Figure 4 (left). Schematic diagram of refrigerator and piping

Figure 5 (right). Layout of the underground piping



house, and a portion of the compressor capacity was released to pick up additional heat from the earth. Because of other physical changes made in the systems, this improvement to the system could not be definitely evaluated, but calculations showed an improvement in efficiency in the order of 5 to 10 per cent.

Air Supply System

The air circulating fan in the Whitemarsh heat pump system was powered by a 1/3-horsepower motor, and it delivered approximately 1,500 cubic feet of air per minute to the house. The fan at Lansdowne was powered by a 1/4-horsepower motor, and it delivered approximately 1,250 cubic feet of air per minute to the house. All of the air was recirculated and filtered at both locations. No outside air was introduced directly into the air supply systems, since infiltration into the houses was sufficient to provide adequate fresh air

Earth Coils and Glycol Circuit

The earth coils at both installations consisted of three loops of pipe connected to a manifold in the basement. The layout of the loops is shown in Figure 5. The 2-inch pipes of the manifold system connected to the loops, which consisted of two $1^1/_2$-inch wrought iron pipes in parallel, 15 inches on center, buried at a depth of 5 feet below the surface. The manifold was so valved that any one, two, or three loops could be used to extract heat from the earth. The Whitemarsh installation contained 2,218 feet of pipe in 1,109 feet of trench, and contained 283 gallons of glycol solution. The Lansdowne installation contained 1,390 feet of pipe in 695 feet of trench, and contained 210 gallons of glycol solution. Each system was equipped with a 1/3-horsepower close-coupled circulating pump to produce a glycol solution flow rate of about 14 gallons per minute. Expansion tanks were provided to take care of the thermal-volume change of the glycol solution, and dirt pockets and strainers were installed to maintain a clean piping system. For the first year's operation, a 17 per cent by volume solution of ethylene glycol was used to protect the system from freezing down to a temperature of 20 degrees Fahrenheit. For the second winter of operation a 22 per cent solution was used to obtain a 12-degree Fahrenheit freezing point. This was done so that information could be obtained regarding the effect of freezing the earth around the coils. Typical illustrations of the pipe installations are shown in Figures 6 to 13.

The Compressors and Unloaders

The hermetically sealed 5-horsepower motor with compressors used, was a standard Chrysler 5-cylinder radial type. However, an unloading device, similar to that used by Chrysler on its larger compressors, was applied to these units. Whenever called upon, compressed air acting through a pneumatic control system operated the unloader to hold open the suction valves on any number of the five cylinders, thus making it impossible for these cylinders to perform work on the refrigerant. The number of cylinders to be unloaded was determined by the outdoor temperature and the temperature of the air supplied to the house. As the outdoor temperature increased, this information was transmitted from an outdoor thermostat to a thermostat in the supply air duct, calling for a lower temperature of the supply air. This thermostat then increased pneumatic pressure on the unloader to position it so that one or more cylinders were cut out of service. On a falling outdoor temperature the reverse action took place to load the cylinders and increase the supply air temperature. When operating in this way the system was known to be on the "improved cycle." The premise was that for this type of operation greater efficiency would result because of decreased power input and lower pressure differentials in the compressor. When operating on the "standard cycle" the compressor ran on all five cylinders at all times when heat was required.

The direct-connected compressor motors operated at 1,750 rpm, and were wired for 3-phase 220-volt 60-cycle current. These units were designed for across-the-line starting.

Figure 6 (left). Whitemarsh — loop A

Figure 8 (right). Loop B in place

AUXILIARY HEATING COILS

Space heaters of the electrical resistance type were installed in the supply air duct of each house. These heaters were used in emergencies only, when the heat pumps were out of service or when they were unable to meet the heating requirements of the homes. The Whitemarsh installation contained 25 kw of resistance heaters in three circuits, one of 5 kw and two of 10 kw. These three circuits were step-controlled by thermostats to provide heat in proportion to the requirements, and high-limit thermostats also were installed. The Lansdowne installation was similarly equipped, except that it consisted of two 5-kw and one 10-kw heating coils. In an emergency these heaters were capable of maintaining the houses at approximately 70 degrees Fahrenheit with zero degrees Fahrenheit outdoors.

CONTROLS

The control system employed standard control equipment available at that time. No specially designed control equipment was used, but a simplified control system especially designed for heat pump use would be employed in production models.

A combination of electric and pneumatic control equipment was furnished with each unit to permit the following operations:

1. Control of the summer cooling cycle from a summer-winter room thermostat. The unloading device was used only as a low limit on the suction pressure.
2. Control of the winter heating cycle by the summer-winter room thermostat, with the following types of operation:
    a. An improved cycle with pneumatic control acting on the unloader to modulate the capacity of the compressor.
    b. Conventional heat pump cycle employing all the cylinders of the compressor while in operation.
3. Provision for operating the fan and the ground coil glycol circulating pump automatically or manually.

Manual and automatic controls also were installed on the emergency space heaters. These controls provided automatic regulation of the space heaters when the heat pump system failed to supply enough heat, and a means of manually controlling the heaters. The glycol pump and fan controls allowed either automatic operation, being cut on and off simultaneously with the compressor motor, or manual operation.

The pneumatic control equipment consisted of an air compressor, tank, and reducing valves to supply air at 15 pounds per square inch for the control circuit, with an additional supply of 30 to 40 pounds per square inch air for compressor unloader control. During the cooling season the compressor unloader was supplied with air through a pressure control which adjusted the unloader to provide an evaporator temperature in the range of 40 to 50 degrees Fahrenheit. In the heating cycle two control sequences were provided. For the improved cycle the unloader was controlled by two thermostats, one located outdoors, and the other located in the air supply which the outdoor thermostat reset in such a manner that the delivered air temperature only slightly exceeded the heating requirements of the building. This is commonly referred to as a "modulated type" of regulation. For the standard cycle the system operated through a pressure control which enabled all five cylinders to function at all times. This usually is called "on-off" control.

Figure 7. Loop B layout

The air pressure within the pressure regulator of the unloader mechanism was varied by these controls. A rise in the air pressure in this assembly reduced the capacity of the compressor until a corresponding rise in the suction pressure of the refrigerant occurred. The pressure change of 15 pounds, the maximum available through conventional control equipment, was not sufficient to give the degree of control required, and therefore a multiplying relay was provided so that a controlled pressure range up to a minimum of 30 to 40 pounds per square inch was available.

Single-pole, double-throw switches were provided for the pump and fan motors. Either switch, when placed in the manual position, caused the corresponding motor to run continuously, whereas, if placed in the automatic position, the corresponding motor operated whenever the compressor operated.

The space heaters controls consisted of a return air thermostat set at approximately 68 degrees and they supplied a series of pressure controls set to energize in sequence the various steps of capacity of the space heater.

TEST EQUIPMENT

A complete testing installation was made on each heat pump unit, providing means to obtain necessary data on the performance of the unit and data pertaining to physical constants involved in heat exchange with the earth.

Electric energy usage was measured by six watt-hour meters as follows:

1. Total input to unit, including compressor motor, glycol pump motor, air com-

Figure 9 (left). Piping to basement

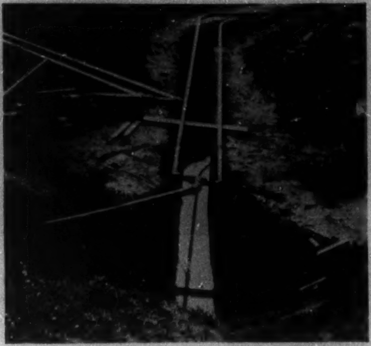


Figure 11 (right). Loop A layout. Note normal ground water

pressor motor for pneumatic controls, air circulating fan motor, emergency space heater, and incidental control board power.

2. Input to main compressor motor operating on heating cycle.
3. Input to main compressor motor operating on cooling cycle.
4. Input to circulating air fan motor.
5. Input to glycol pump motor.
6. Input to emergency space heater.

Energy used by the control equipment, including the air compressor motor for pneumatic controls, was obtained by the proper difference between the foregoing data.

Demand characteristics of the unit were determined by three type-*G9* demand meters which showed integrated 30-minute demands for:

1. The compressor motor when operating on the heating cycle.
2. The compressor motor when operating on the cooling cycle.
3. The emergency space heater when used.

Figure 10. Excavation at Lansdowne

As an added check on the actual hours of operation of the compressor motor, two running time meters were installed, one for operation on the heating cycle, the other for operation on the cooling cycle.

An Esterline graphic wattmeter was installed on the supply to the compressor motor. An integrating water meter was installed in the ground coil system to measure glycol flow through the pipes. Temperature measurements throughout the system were made with thermocouples. A 16-point Micromax temperature recorder kept a graphic record of the temperatures at the selected points of the system. Graphic records were kept of the temperature differential of the circulating glycol solution to and from the ground coils in use, by means of hot and cold thermohms.

Temperature measurements at various locations in the earth near the ground coils were taken by thermocouples connected to selector switches. These temperatures were recorded at the time of periodic visits by a laboratory technician.[1]

All instruments and recorders for the test equipment were located on the test panels in the vicinity of the actual unit. Weekly visits were made by a laboratory technician to obtain operation data and to adjust the instruments. The time between visits was increased to biweekly and monthly in the interest of economy when it became apparent that sufficient data could be obtained with less attention to the equipment. No data, except kilowatt-hour usage, were taken during the second summer, since it was felt that sufficient information had been obtained on the cooling operation during the first year.

Several thermocouples were relocated on the Micromax recorder for the second year's operation when certain points were no longer of any particular value, and other temperatures were known to be of greater importance. Thermocouple wiring during installation is shown in Figure 14.

Figure 12. Loop A in place showing test connection

Figure 13 (left). Quicksand at Lansdowne

Figure 15 (right). Piping manifold at Lansdowne

Figure 15 shows the piping manifold, and Figure 16 is a view of the control and power panels.

Performance

Essential data of results obtained are shown in graphic form in Figure 17 for ease of interpretation. In comparing the results at the two locations, certain variations can be attributed to differences in heat demand of the houses, differences in occupancy and use, and variation in local weather conditions.

Ambient Earth Temperatures

The ambient earth mean temperatures for the test periods at the 5-foot coil depth show the seasonal variation observed at the two test locations, and the variations that may be expected in this area for this type of heat source. The high ambient of 65 degrees Fahrenheit in the fall dropped to a low of 45 degrees Fahrenheit by early spring. Study of investigations made by the United States Bureau of Mines in 1941 assisted in the decision to install the earth coils at the 5-foot depth.

Ground Coil Temperature

The average ground coil temperatures were obtained from recorded data of thermocouples located on the pipe at the midpoint of loop A. They show an increasing differential below the ambient earth temperature as the weather became colder and as the total quantity and rate of heat withdrawal increased. The differentials reached a stablized condition after about 2 months of heat pump operation, with the coil operating at a temperature of 15 to 20 degrees below the earth ambient.

Evaporator Temperatures

During the first heating season, there was considerable fluctuation in the average refrigerant temperature in the glycol heat exchanger during certain test periods. The variations may be attributed to changing from the standard cycle to the improved cycle. The refrigerant temperatures were, on the average, approximately 7 degrees Fahrenheit below the ground coil temperatures, dropping as the heat demand increased. During the second season, a more regular differential was obtained when the improved cycle was abandoned as not showing any improvement in performance of the systems. The marked decrease in the Whitemarsh refrigerant temperatures in the second season was due to increasing the capacity of the system from 5 horsepower to 7½ horsepower to match more nearly the heating load without changing the size of the piping or heat exchangers. The large increase in differential below the earth coil temperature in the latter half of the second season was due to an attempt to freeze the earth surrounding the earth coil. The low limits of the compressor suction temperature controls were lowered and additional ethylene glycol was added to the earth coil solution to permit safe operation at the lower temperatures.

Length of Ground Coil

The length of the ground coil in operation as a heat source was changed from time to time during the first season. No appreciable effect on performance was noted between test periods in which minimum and maximum lengths of coil

Figure 14 (below). Thermocouple wire installed at Whitemarsh

Figure 16 (right). Control and power panels at Lansdowne

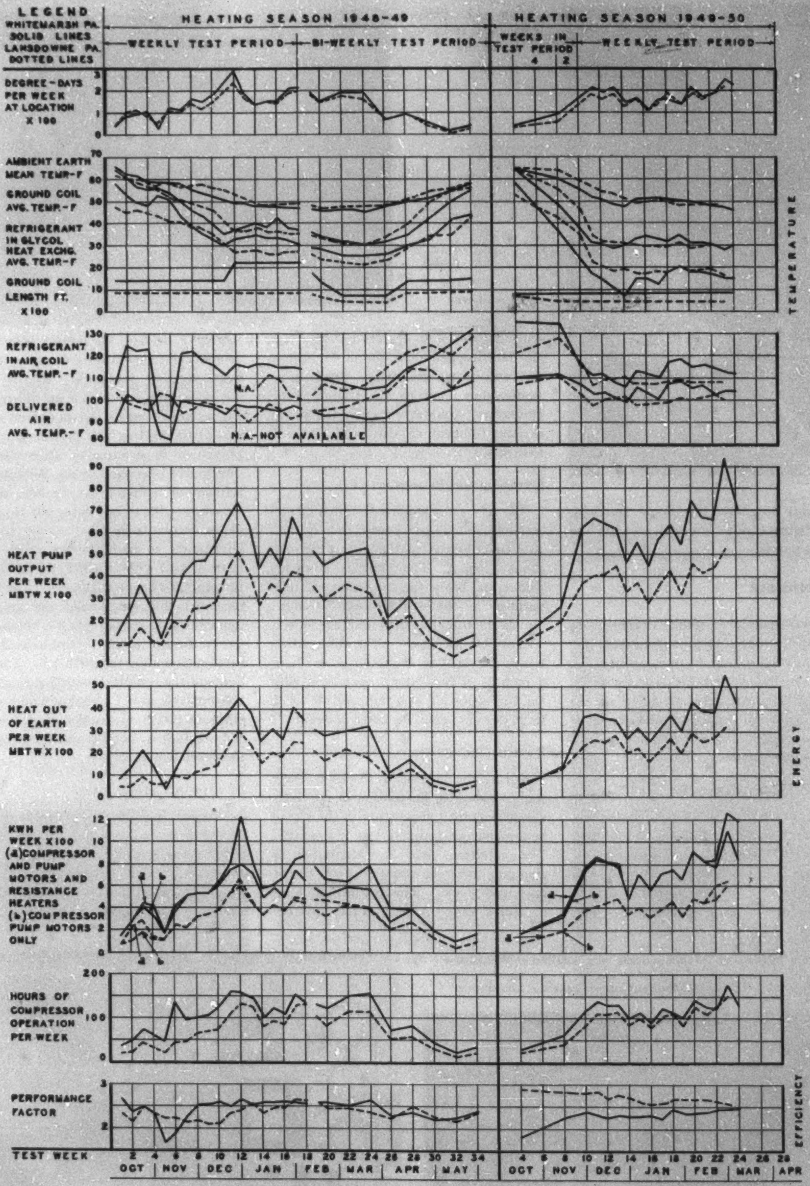


Figure 17. Graphic presentation of heat pump performance

were used. It became obvious that the amount of coil in one loop was adequate to meet the needs of the installed compressor capacities. Further details on this subject are presented in the paper dealing with earth as a heat source.[1]

CONDENSING TEMPERATURES

The average temperatures of the refrigerant in the air coil showed wide fluctuation during the greater part of the first season, which is attributed to alternate operation on the standard and improved cycles, as well as to weather changes. Lansdowne data were not available for a period because of a broken thermocouple. During the periods of severest winter weather these coil temperatures averaged 110 degrees Fahrenheit.

DELIVERED AIR TEMPERATURES

The average temperatures of the air delivered from the coil generally ranged from 7 to 15 degrees below the condensing coil temperatures, and averaged 10 degrees below those temperatures. However, no discomfort was reported by the occupants of the homes with delivered air temperatures of 100 degrees Fahrenheit, or even temperatures as low as 90 degrees Fahrenheit. If anything, comfort was improved at these temperatures because the heat pumps ran more continuously, giving an even flow of air at constant temperatures. In the normal air heating systems, air as hot as 140 degrees Fahrenheit is supplied for short periods only, creating discomfort due to wide changes in air temperature and motion.

ENERGY OUTPUT AND INPUT

The next set of curves shows the relation between the weekly output of the systems, the heat removed from the earth for that period, and the electric input to the systems. The heat pump outputs for the 5-horsepower units averaged about 40,000 Btu per hour, while the rates of heat withdrawal from the earth were about 25,000 Btu per hour. This indicated that about 15,000 Btu per hour of the heat supplied were furnished by the power input to the systems, with a performance factor of approximately 2.6. The $7^1/_2$-horsepower unit installed at Whitemarsh on November 10, 1949, gave a higher output and rate of output in proportion to the change in horsepower. The weekly output and input curves show a normal trend corresponding to the trend of degree days. The requirement for extra heat by the use of the emergency electric resistance heaters coincided closely with the severity of weather. At Whitemarsh the use of supplemental heat was greater than at Lansdowne, because the heat pump unit capacity did not approach as closely the heating requirements of the house. Also, at Whitemarsh the number of degree days of heating requirement was more than 10 per cent greater than at Lansdowne. It will be noted that during the second season the amounts of supplemental heat at both locations were reduced considerably below the amounts in use during the first season because of increased compressor capacity at Whitemarsh and improved compressor efficiency at Lansdowne.

HOURS OF OPERATION

The curves of hours of compressor operation per week show the more continuous operation of the heat pump as compared to conventional heating systems. Periods of almost continuous operation for the severest weather may be noted. The house thermostats were maintained at 70 to 72 degrees Fahrenheit continually throughout the heating season. No night temperature cutback was attempted because the limited capacity of the units would have made it impossible to obtain rapid temperature pickup in the morning. This, of necessity, would become standard heat pump operating practice, since it would not be economically feasible to install the required extra capacity necessary for quick pickup.

PERFORMANCE FACTORS (COP)

Performance factors remained relatively constant throughout the heating season. During the first season, the average at both locations was approximately 2.5. No difference was noted in the final results on improved as compared with standard cycle operation. This may be attributed to the fact that any power savings made in the improved cycle due to partial loading of the cylinders, and operation with smaller pressure differentials, was largely offset by the continued friction and windage losses in the unloaded cylinders. As a result, the units were operated only on the standard or 100-per cent on-off cycle for the second season. The Lansdowne installation showed an improvement in performance factor to an average of 2.7. Prior to the second heating season, a new motor-compressor had been installed at this location. Improved valve and piston design in the new compressor may be credited with the better performance. At Whitemarsh, at the beginning of the second season, the poor performance was caused by improper refrigerant charge and control calibration. Corrective measures were not taken immediately because it was anticipated that the new compressor would be installed shortly. However, unforeseen delays in shipment of the compressor prolonged this period. After installation of the $7^1/_2$-horsepower compressor, the performance at Whitemarsh improved considerably. Two factors affected the performance to keep it below that of the Lansdowne installation. First, the increased capacity of the compressor was obtained by increasing the length of the piston stroke by welding sections in the connecting rods, installing new head castings, and installing an external discharge header. These changes were hurriedly designed and made, and did not represent the ultimate in efficiency of performance. Second, the capacity of the air heating coil was increased. Had the capacity of the heat exchangers, piping, and other equipment been increased, better efficiency undoubtedly would have been achieved. Lack of time and space in the existing units precluded making these other changes.

SUMMER COOLING PERFORMANCE

During the first summer, data were taken to determine the performance on the cooling cycle, but the information was not analyzed completely. Actual performance proved to be very satisfactory, and it was felt that analysis would not provide sufficient information to justify the time consumed. Relatively constant room thermostat settings were maintained, and the units proved quite adequate in handling the cooling load. No difficulties were encountered which could be attributed to cooling operation, other than the fact that additional Freon was required in the refrigerant circuit for most efficient operation on cooling. No corrective measures were taken on this design problem, which arose from the physical layout of the equipment and the fact that operating pressure ranges differed with cooling or heating service. A negligible amount of data was taken during the second summer.

SUMMARY OF DATA

A summary of significant data for the two test year periods is shown in Table I for the two installations. Attention is directed to the following important features contained in this table as a result of data obtained from the tests.

1. The annual load factors of the heat pump system, with the use of supplementary resistance heating elements, were exceedingly low.
2. Even under a restatement of results, for

Table I. Summary of Significant Data for Heat Pump Installations at Whitemarsh and Lansdowne for 1948–49 and 1949–50 Test Periods

	Whitemarsh		Lansdowne	
	1948–49	1949–50	1948–49	1949–50
Calculated heating requirements under standard conditions (0 degrees Fahrenheit outdoors, 70 degrees Fahrenheit indoors)—Btu per hour	90,000	90,000	75,000	75,000
Rated capacity of heat pump compressor, horsepower	5	7.5	5	5
Operating experience:				
1. Number of days in test period	350	354	351	347
2. Total heating season degree-days at location	4,459	4,750	4,006	4,235
3. Maximum output of heat pump during December, January, and February, Btu per hour	45,700	55,100	38,300	38,800
4. Ratio of maximum output of heat per hour during December, January, and February to calculated heating requirements, per cent	50.8	61.2	51.1	51.7
5. Length of ground coil in use during the maximum output of heat pump, feet	2,220	770	860	450
6. Ratio of compressor operation during December, January, and February to total hours in this period, per cent	77.3	72.3	58.1	63.6
7. Performance factor for the heating season	2.49	2.32	2.38	2.71
8. Electric load characteristics*				
A. Annual energy use, kilowatt-hours				
Heating:				
Heat pump (including Glycol pump)	15,248	20,835	9,855	10,801
Resistance elements	2,384	772	538	732
Total for heating	17,632	21,607	10,393	11,533
Cooling	3,491	2,796	1,204	705
Total heating and cooling	21,123	24,403	11,597	12,238
B. Maximum 30-minute demands:				
Annual of system (including resistance elements)	21.4	21.9	20.8	19.8
Heat pump (including Glycol pump) during December, January, and February	5.4	6.7	5.1	4.3
C. Annual load factors, per cent:				
Based on annual maximum demand of system:				
Heating only	9.4	11.3	5.7	6.7
Total heating and cooling	11.3	12.7	6.4	7.1
Based on heat pump maximum demand during December, January, and February and excluding energy use of resistance elements	39.6	40.3	24.8	30.5
D. Total energy use for heating per degree-day	3.96	4.55	2.60	2.72
Restated operating experience†				
1. Electric load characteristics				
A. Annual energy use, kilowatt-hours				
Heating	16,205	21,168	10,081	11,071
Cooling	3,491	2,796	1,204	705
Total heating and cooling	19,696	23,964	11,285	11,776
B. Annual maximum 30-minute demand of system, kilowatts	10.6	11.4	9.2	8.1
C. Annual load factors, per cent:				
Heating only	17.5	21.2	12.5	15.6
Total heating and cooling	21.2	24.0	14.0	16.6
D. Total energy use for heating per degree-day	3.63	4.46	2.52	2.61

* Annual energy use and factors dependent on it reflect test period experience adjusted to 365 days.

† Restated operating experience reflects estimated conditions, which would have existed, under the actually experienced performance factors, had the maximum output of the heat pump (during the winter period) been equal to the calculated heating requirements.

conditions which would have been experienced had the maximum output of the heat pump units (during the winter period) been equal to the calculated design heating requirements of the premises, the annual load factors of the heat pump system would not have reached satisfactory levels from the standpoint of the economic significance of electricity supply.

3. The annual load factors obtained by the heat pump units themselves, however, reached desirable levels largely because the heating capacity of these units was considerably undersized in relation to the calculated design heating requirements of the premises. This desirable effect on the annual load factor, in turn, indicates the necessity for providing a heating capacity of heat pump units considerably below the design heat requirements of the premises, provided the supplementary requirements for heating, which would be required during protracted cold spells, is introduced in such a way as not to exceed the maximum demand of the heat pump unit on the coldest day.

4. Heat pump units which are undersized in relation to the heating requirements of the premises will be subject to long periods of operation in a year, especially during the cold winter months. This imposes severe requirements upon their design and quality of construction, as well as upon the design of the heat transfer surfaces from and to the earth, when that medium is used as a heat source or a heat sink.

Difficulties in Operation

The major trouble sources of these two installations can be classified in two categories: first, an abnormal number of Freon leaks were experienced; and second, difficulties arose due to the complex controls and instrumentation used on the units. Several mechanical breakdowns or faults also were experienced during the test period.

REFRIGERANT LEAKS

Approximately 12 Freon leaks occurred on the two units, and each of these was found to be the result of poor workmanship, such as a number of improperly flared connections, several screwed connections with only one or two threads engaged, and flanged connections with only a partial bolt thread engaged.

Failures of this type were not anticipated, since they are normally very rare in package-type air-conditioning units. They occurred in these experimental units because a number of special parts were used in the heat pumps and they were not subjected to the careful inspection procedure of a production line. It is not anticipated that this type of trouble will be encountered in the future on production line models.

CONTROLS

Use of a complex system of pneumatic and electric controls added to the difficulties. The combination of controls did not prove to be a happy one, but the use of the compressor unloading device, as well as change-over provisions from improved to standard cycle operation, dictated their use. It was very difficult to achieve proper setting of the complete control system, and it also proved to be difficult to maintain proper calibration. During the first season, constant readjustment was required to maintain the proper relationship between controls as the temperature ranges of outside air, earth, refrigerant, and supply air changed. Abandonment of the improved cycle of operation for the second heating season made possible some simplification of controls. Plans had been completed to reduce controls to a minimum during the third season, but the units were removed and the experiment concluded before this change was made. In a production line heat pump unit, simple, basic electric controls would be adequate, and no difficulties should be encountered from this source.

INSTRUMENTS

Instruments, while not adversely affecting the performance of the units, did create some problems of their own. Extreme sensitivity, particularly of the differential temperature recorder, required that the glycol piping exposed in the basements be insulated. The thermohm elements for this instrument were so sensitive that they had to be insulated and shielded from all surrounding sources of radiant heat, no matter how minute. Thermocouples placed in various locations in some instances also required special insulation or relocation to provide accurate data. These difficulties, of course, were due only to the experimental nature of the installations.

MECHANICAL FAULTS

Several mechanical faults encountered during the period of operation are noted below.

1. The fan motor at Whitemarsh burned out and was replaced. This motor had been used for a period of 10 years in the original gas-fired furnace at this location, and was reused to power the air-circulating fan for the heat pump. This can be expected to occur in normal operation for the indicated life span of the equipment.

2. The variable pitch fan motor pulley at Whitemarsh opened up because of a loose-set screw, and belt slippage occurred. Also on one occasion the belt stretched and slippage occurred.

3. Tests indicated excess clearances in the automatic 4-way refrigerant valves at Lansdowne, causing a by-passing of some of the refrigerant with resulting reduction in efficiency of the unit. A new valve cage was installed to correct this condition.

4. During the first summer, the main bearings of the compressor at Whitemarsh seized, as a result of loss of oil from the crankcase, requiring replacement of all bearing inserts. General engineering opinion indicates that this loss of oil was due to liquid refrigerant returning to the crankcase, causing foaming of the oil and carrying it over into the rest of the refrigerant system.

Upon completion of the second heating season, it was decided that the units be allowed to run, without special attention, to satisfy the heating and cooling requirements of the homes until a serious breakdown occurred. This was done to determine the life expectancy of the units.

Early in November 1950, the Lansdowne compressor failed as a result of loss of oil in the crankcase, similar to the previous failure at Whitemarsh. The unit was removed and replaced with a gas-fired furnace. This type of failure is not common in our experience with numerous air-conditioning installations.

The Whitemarsh unit failed during the severe storm which occurred over the Thanksgiving week end of 1950. Analysis of available facts indicates that the following sequence of events took place. A falling tree a short distance from the Whitemarsh residence evidently tore down one wire of the 3-phase line installed to supply the heat pump compressor motor. Failure of phase protection to operate caused the motor to run on single phase and burned out one winding. All power to the home was lost then, and when service was re-established the motor failure was discovered. The unit was replaced with a gas-fired furnace.

Conclusions

The experience gained by the company on the installation and operation of two experimental heat pumps to provide year-round air comfort in the home leads to the following conclusions.

1. The two experimental heat pump installations performed effectively their task of providing year-round air comfort in the home.

2. With due allowance for the experimental nature of the installations, the occupants of both houses expressed full satisfaction on the heating and cooling effects of their heat pump installations.

3. The earth provided a reliable heat source and heat sink during the two years of heat pump operation, including the periods of the highest heating and cooling requirements.[1]

4. Substantial improvements must be made in the mechanical design, construction, and assembly of heat pump units if they are to obtain wide commercial acceptance as to the reliability of their operation and economy of performance.

5. To obtain maximum desirability as a builder of utility load under the most advantageous competitive conditions with other fuels, it is desirable that the heating capacity of heat pump installations be limited to their capacity for cooling requirements, and supplementary heating requirements must be met in such a manner as not to create demands on the supply system which would exceed the maximum demand of the heat pump unit during the coldest days of the heating season.

6. The foregoing improvements might be achieved through the introduction of heat storage with the heat pumps. At present, progress is reported on investigations being conducted in this country on the use of low-cost material which would store heat in a minimum of space. As applied to the heat pump, this will make possible the use of units with smaller capacities, electrically operating under considerably improved annual load factors. This feature will make the heat pump load economically attractive to electrical utilities.[2]

References

1. RESIDENTIAL HEAT PUMP EXPERIMENTS IN PHILADELPHIA—EARTH AS A HEAT SOURCE, A. H. Kidder, J. H. Neher. *AIEE Transactions*, volume 71, part II, 1952, pages 343–50.

2. RESIDENTIAL HEAT PUMP EXPERIMENTS IN PHILADELPHIA—SUGGESTED POSSIBILITIES FOR PRACTICAL APPLICATIONS, Constantine Bary. *AIEE Transactions*, volume 71, part II, 1952 (January 1953 section), pages 361–65.

No Discussion

Continuous Current and Temperature Rise in Bundled Cables for Aircraft

MILTON SCHACH
ASSOCIATE MEMBER AIEE

R. E. KIDWELL, JR.
ASSOCIATE MEMBER AIEE

The need for bundle ratings arises from the fact that cables installed in aircraft are grouped together for a portion of the cable run from the distribution bus to the individual load devices. Stated in its simplest form, the bundle-rating problem is to determine a method for selecting the proper combination of cable sizes for any given group of currents; or stated, conversely, given a group of cables, to find the combination of currents which may be carried safely by the group.

There are two general criteria by which the current capacity of a cable is limited: 1. The maximum voltage drop in the conductor. 2. The maximum insulation temperature. This paper will be concerned with the latter only. The fact that the conductor temperature is nearly uniform over its cross section and the fact that the temperatures of the insulation and the conductor approach each other at their common boundary make it possible to state the thermal limitation on current capacity as a maximum conductor temperature. When applied to a cable bundle, this criterion requires that the hottest conductor must not exceed some predetermined temperature.

The factors which determine the temperature of the hottest conductor are the number and properties of the component cables, the number of cables loaded simultaneously, the position of the loaded cables in the bundle, the over-all size and shape of the bundle, the spacing between adjacent cables, and the bundle environment.

Present ratings for cables in bundles are based on measurements conducted by the United States Bureau of Standards in 1942. These measurements determined the current per cable necessary to produce a 42.8-degree-centigrade rise in conductor temperature for three like cables in conduit when the three cables are loaded simultaneously. The conduit was located in air at an ambient temperature of 57.2 degrees centigrade and at sea-level pressure. The need for a method of rating cables based on conditions which conform more closely to actual installations and which permits greater flexibility in loading is generally acknowledged.

The general solution of the bundle-rating problem is not immediately attainable for four reasons. First, there is a lack of information concerning the numerous factors. Second, it is difficult to express quantitatively such parameters as bundle shape and spacing between adjacent cables. Third, the criterion that the hottest conductor in a bundle must not exceed some specified temperature permits many solutions for the problem, that is, by decreasing the current in some cables and increasing it in others, one obtains alternative current values which satisfy the criterion. Finally, the complex bundle geometry gives rise to elaborate boundary conditions for conduction, convection, and radiation.

Simplification of Problem

Recognition of the difficulties of the bundle-rating problem in its most general form made it evident that an investigation of the problem under much more restricted conditions than those which prevail in practice should be undertaken. In particular, it was believed that an empirical study of uniform bundles (bundles composed of cables of a single size) would be useful not only in providing data on the direct effects of several of the factors affecting current rating but also in determining whether an analysis of the heat flow within and from the bundle was possible. The study was conducted under the following simplifying conditions:

1. Cables were bound at intervals not exceeding 24 inches so as to form a cylindrical-shaped bundle.
2. The relative positions of the component cables were maintained throughout the bundle length.
3. Bundles were suspended horizontally in still air. For convenience, the ambient air was fixed at room temperature (25 degrees centigrade) and sea-level pressure.
4. Each loaded cable carried the same current.

Definition of Current Rating

For convenience, current rating is defined as the continuous current per loaded cable which produces a 40-degree-centigrade rise in the temperature of the hottest conductor of the bundle. The terms "current capacity," "current per cable," and "current value" will be used synonymously with "current rating" as defined here. The selection of 40 degrees centigrade as the maximum temperature rise is arbitrary. Where values other than 40 degrees centigrade are discussed, they will be indicated specifically.

Factors Affecting Current Rating

With the restrictions mentioned, current rating becomes a function of the dimensions, resistance, and thermal properties of the component cables, the total number of cables in the bundle, the number of cables loaded simultaneously, the position of the loaded cables in the bundle, and the spacing between adjacent cables.

Experimental Procedure

The measurements were initiated with a bundle of 19 *AN-18* (Army-Navy) cables bound tightly at 1-inch intervals in a manner to minimize the distance between adjacent cables (bundles tied in such a manner are called "tight" bundles). A diagram of the cross section of such a bundle of 19 cables is shown in Figure 1. A series of tests was performed to determine current rating as a function of the position of the loaded cables and of the number of cables loaded simultaneously. Only those positions in which the loaded cables were situated symmetrically about the bundle center and those which gave either the best or the poorest current rating were chosen.

To determine the effect of the total number of cables in the bundle on current rating, tests similar to those made on the initial bundle were performed on tight bundles of 3, 7, and 37 *AN-18* cables. In addition, the current rating was obtained for a single *AN-18* cable in air.

To ascertain the effect of the spacing between adjacent cables on current rating, tests similar to those conducted on tight bundles were performed on a bundle of 37 *AN-18* cables tied at 6-, 12-, and 24-inch intervals. (Bundles tied at 6-, 12-, or 24-inch intervals are denoted "loose" bundles.)

The whole series of tests was repeated on bundles of 3, 7, and 19 cables of size

---

Paper 52-323, recommended by the AIEE Air Transportation Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Middle Eastern District Meeting, Toledo, Ohio, October 28-30, 1952. Manuscript submitted May 5, 1952; made available for printing September 5, 1952.

MILTON SCHACH and R. E. KIDWELL, JR., are both with the Naval Research Laboratory, Washington, D. C.

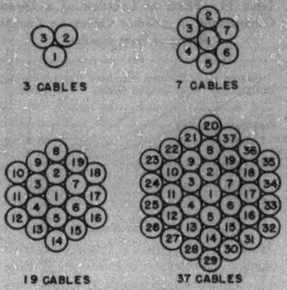


Figure 1. Cross sections of tight bundles of 3, 7, 19, and 37 cables showing the cable position numbers

AN-8 and of size *AN-1/0* in order to determine the effect of the size of the constituent cables on current rating. In addition, current ratings for a single cable in air were obtained for these sizes.

A complete list of the bundles studied is given in Table I; bundles will be referred to by the designations given there. The numbers which make up the designation are the total number of cables in the bundle, the individual cable size, and the interval between ties, in that order. Diagrams of the tight bundles are shown in Figure 1; the numbers which appear on the diagram are given so that cable positions may be located. The manner of binding tight and loose bundles is illustrated in Figure 2 for the case of 19 *AN-8* cables. In tying the tight bundles, lacing of 1-inch interval was used in all cases except for bundle 19-1/0-3, where lacing of 3-inch interval was deemed sufficient to produce a tight bundle. A description of the methods employed and the conditions under which measurements were obtained is given in Appendix I.

Empirical Results (Tight Bundles)

In the following description of the empirical results, each factor affecting current rating is considered in turn. When the effect of one factor is being discussed, it may be assumed that those remaining are held constant except where specifically indicated. Emphasis is placed on the conditions for which minimum current ratings are obtained.

NUMBER OF CABLES LOADED

If every possible combination of cables in a given bundle were loaded in turn and the current per cable for each test were plotted as a function of the number of cables loaded, it would be found that current rating is not a single-valued function. This result is illustrated for a tight bundle of 37 *AN-18* cables in Figure 3. Points which lie on the upper and lower curves correspond to the best and poorest heat transfer conditions respectively. The spread of current values is due to the variety of positions which can be occupied by a given number of loaded cables. Additional information concerning the effect of the number of cables loaded is presented as the other factors are introduced.

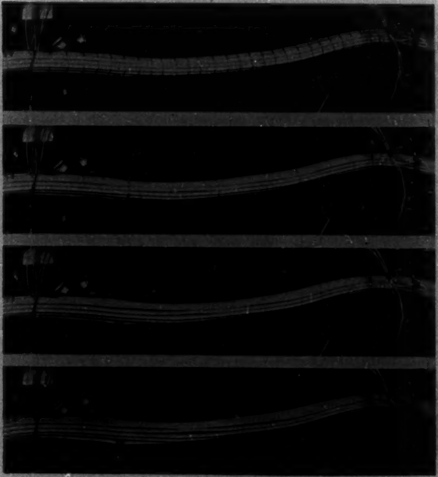


Figure 2 (left). Bundles of 19 AN-8 cables bound by 1-inch lacing and by 6-, 12-, and 24-inch ties

Table I. List of Bundles Studied

Bundle Designation	Number of Cables in Bundle	Army-Navy Size of Constituent Cables	Distance between Ties, Inches
3-18-1	3	18	1
7-18-1	7	18	1
19-18-1	19	18	1
37-18-1	37	18	1
37-18-6	37	18	6
37-18-12	37	18	12
37-18-24	37	18	24
3-8-1	3	8	1
7-8-1	7	8	1
19-8-1	19	8	1
19-8-6	19	8	6
19-8-12	19	8	12
19-8-24	19	8	24
3-1/0-1	3	1/0	1
7-1/0-1	7	1/0	1
19-1/0-3	19	1/0	3
19-1/0-6	19	1/0	6
19-1/0-24	19	1/0	24

POSITION OF THE LOADED CABLES

The term position may be used in two senses. First, it may refer to the configuration of the loaded cables, that is, the pattern of and distances between the current-carrying cables. Second, it may mean the location of a given configuration of loaded cables with respect to the bundle center.

Configuration of the Loaded Cables. The current capacity increases as the distance between loaded cables is increased. Current capacity is a minimum when the loaded cables occupy adjacent positions in the bundle and form a compact group. The ratio of current ratings for the two configurations which give maximum and minimum values for the same number of cables loaded simultaneously can be as high as 1.3. A summary of the configuration effect is given in Table II.

Location of a Given Configuration of Loaded Cables with Respect to the Bundle Center. The location for which a given configuration has a minimum current rating varies with the number of cables loaded and the total number of cables in the bundle. If the current capacity of the

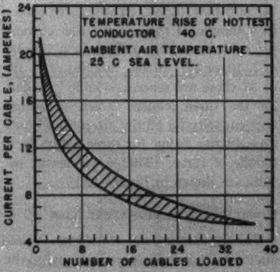


Figure 3 (right). Effect of the number of cables loaded on current rating for a tight bundle of 37 AN-18 cables

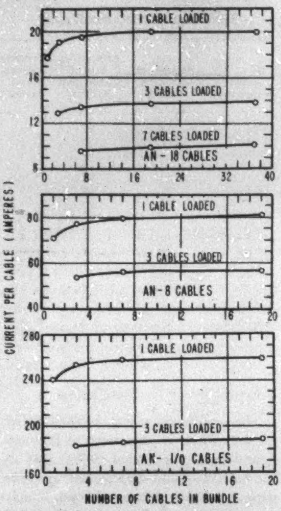


Figure 4. Effect of the total number of cables in the bundle on current rating. Each point plotted represents the lowest current rating for the specified number loaded. Temperature rise of the hottest conductor, 40 degrees centigrade; ambient air temperature, 25 degrees centigrade; sea level

centermost position is taken as a reference, it is found that the current values for alternative locations differ by less than 7 per cent in each of the tight bundles. The effect of location of the loaded cables on current rating is summarized in Table III.

Total Number of Cables in the Bundle

Over the range observed, current rating for a given number of cables loaded increases as the total number of cables increases. In particular, current capacity is a minimum when the total number of cables in the bundle is equal to the number of cables loaded. Thus, three cables in a bundle of three carry less current per cable than three cables in a bundle of 7, 19, or 37. In the bundles of 7, 19, or 37 there are several possible locations for the three loaded cables. The current capacity for all the alternative cases exceeds the value for three cables in a bundle of three. The experimental results are shown in Figure 4.

Current and Temperature Rise

The temperature rise of the hottest conductor in the bundle is proportional to the square of the current per loaded cable for all fully loaded, tight bundles over the range of temperature rise investigated (approximately 20 to 80 degrees centigrade). Temperature rise values calculated with this relation agreed with measured values to within 3 per cent. The experimentally determined current and temperature rise for fully loaded bundles of sizes *AN-18*, *AN-8*, and *AN-1/0* are shown in Figures 5, 6, and 7.

Table II. Effect of Configuration of Loaded Cables on Current Rating of Tight Bundles

Bundle Designation	Position Numbers of Loaded Cables	Current Rating, Amperes
Three Cables Loaded		
19-18-1	2, 8, 9	13.8
	8, 9, 19	14.0
	8, 12, 16	16.3
37-18-1	8, 20, 37	14.3
	20, 26, 32	16.6
Six Cables Loaded		
19-18-1	1-3, 8-10	10.8
	2-7	11.2
	9, 11, 13, 15, 17, 19	12.5
37-18-1	1-3, 8-10	10.8
	20, 23, 26, 29, 32, 35	13.9
19-8-1	1-3, 8-10	44.5
	2-7	46.0
19-1/0-3	1-3, 8-10	149
	2-7	152
Twelve Cables Loaded		
19-18-1	1-3, 5, 7-11, 17-19	8.2
	8-19	9.1
37-18-1	1-3, 7-10, 19-22, 37	8.4
	8-19	9.1
	9, 11, 13, 15, 17, 19, 20, 23, 26, 29, 32, 35	9.7
	2, 4, 6, and 21, 23, 25, 27, 29, 31, 33, 35, 37	9.7
	20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37	10.4
19-8-1	1-3, 5, 7-11, 17-19	35.5
	8-19	39.5
19-1/0-3	1-3, 5, 7-11, 17-19	118
	8-19	138

Summary

It has been noted that current capacity is not a single-valued function of the number of cables loaded and that the multiple values result from the fact that there is more than one position for each given number of cables loaded. A single-valued function can be obtained for any given bundle by considering only those positions (corresponding to poorest heat transfer conditions) which give the lowest ratings. Such experimentally determined curves, marked (1), are shown in Figures 8, 9, and 10 for tight bundles 37-18-1, 19-8-1, and 19-1/0-3.

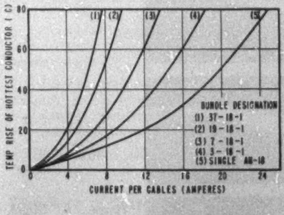


Figure 5. Temperature rise of the hottest conductor as a function of current per cable for tight, fully loaded, uniform bundles of AN-18 cables. Ambient air temperature, 25 degrees centigrade, sea level

Table III. Effect of Location of a Given Configuration of Loaded Cables on Current Rating of Tight Bundles

Bundle Designation	Position Numbers of Loaded Cables	Current Rating, Amperes
One Cable Loaded		
7-18-1	1	20.1
	2	19.6
19-18-1	1	21.0
	2	20.6
	8	20.0
37-18-1	1	21.2
	20	20.1
7-8-1	1	83.3
	2	79.7
19-8-1	1	84.0
	8	80.0
7-1/0-1	1	271
	2	264
	5	257
19-1/0-3	1	272
	8	262
	14	259
Three Cables Loaded		
19-18-1	1, 2, 3	14.1
	2, 8, 9	13.8
37-18-1	1, 2, 3	13.9
	2, 8, 9	13.9
	8, 20, 37	14.3
19-8-1	1, 2, 3	58.0
	2, 8, 9	56.5
19-1/0-3	1, 2, 3	189
	2, 8, 9	189
	5, 13, 14	189
Six Cables Loaded		
37-18-1	2, 8, 9, 20-22	11.3
	1-3, 8-10	10.8

It was shown that the current rating for a given number of cables loaded is a minimum when the total number of cables in the bundle equals the number loaded. If current rating based on fully loaded bundles (bundles in which every cable carries current) is plotted as a function of the number of cables loaded, the minimum curve for the tight-bundle condition

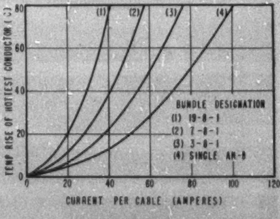


Figure 6. Temperature rise of the hottest conductor as a function of the current per cable for tight, fully loaded, uniform bundles of AN-8 cables. Ambient air temperature, 25 degrees centigrade; sea level

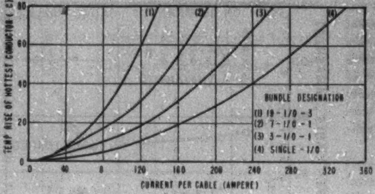


Figure 7 (left). Temperature rise of the hottest conductor as a function of the current per cable for tight, fully loaded, uniform bundles of AN-1/0 cables. Ambient air temperature, 25 degrees centigrade; sea level

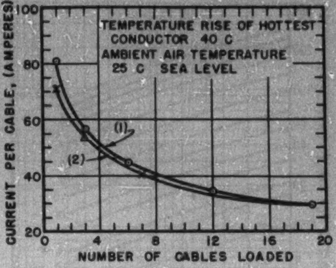


Figure 9 (right). (1) Lowest current ratings for a tight bundle of 19 AN-8 cables. (2) Current ratings for fully loaded, tight bundles of AN-8 cables

is obtained. Such curves, marked (2), are shown in Figures 8, 9, and 10 for *A N* cable sizes 18, 8, and 1/0 respectively. These curves may be described by the approximate relation

$$I_n = I_1 / n^{\lambda}$$

where I_1 is the current which may be carried by a single cable in air, n is the number of cables in the fully loaded bundle and I_n is the current per cable for the n cables. The exponent λ has the values 0.324, 0.298, and 0.299 for bundles of *AN-18*, *AN-8*, and *AN-1/0* cables respectively. Minimum current ratings for tight bundles of typical aircraft cables are given in Figure 11; the curves for *AN* sizes *18*, *8*, and *1/0* are based on actual measurements, the remaining curves being interpolated from these. The single-cable current values used in the interpolated curves are based on Military Specification *MIL-W-5088*.

Empirical Results (Loose Bundles)

As the interval between ties is increased, that is, as the bundle is loosened, current capacity changes gradually. The magnitude and sign of the change depends on the number and position of the loaded cables. The experimental results for bundles of 37 *AN-18*, 19 *AN-8*, and 19 *AN-1/0* cables are shown in Figures 12, 13, and 14. The position numbers of the loaded cables are indicated for each curve. It may be noted that the interval of tie for which a minimum current capacity is obtained depends upon the number of cables loaded. Current values for the poorest heat-transfer conditions are shown as a function of the number of cables loaded in Figures 15, 16, and 17 for loose bundles of 37 *AN-18*, 19 *AN-8*, and 19 *AN-1/0* cables by curves marked (2).

The changes in current capacity which occur when the bundle is loosened may be explained qualitatively in terms of the modified heat-transfer conditions. Two opposing processes occur as the bundle is loosened: 1. The effective thermal conductivity of the bundle decreases. 2. The conditions for heat transfer by convection and radiation at the outer bundle surface improve, that is, the effective surface area increases. For the case of one cable at the center loaded, the deterioration in heat conduction is the predominant effect, and the current rating is reduced when the bundle is loosened. For the case in which the entire outer layer and no other cables carry current, the only effect to be considered is that of the improved heat transmission at the outer surface. Consequently, the current rating for this case increases as the bundle is loosened. Finally, if every cable in the bundle carries current, the two effects nearly compensate each other, and so the current capacity is nearly independent of bundle tightness for the range investigated. In all cases, if the cables were to be separated further than observed experimentally, the current in each loaded cable would begin to increase and approach its value for a single cable in air.

Tight Versus Loose Bundles

Figures 15, 16, and 17 compare the minimum tight- and loose-bundle ratings. The loose-bundle conditions give the absolute minimum current rating for each number of cables loaded. It was observed that the minimum values for bundles bound by 6-inch ties are as much as 3 per cent below the tight-bundle minimum values while the lowest values of bundles bound by 24-inch ties are as much as 6 per cent below the minimum tight-bundle values.

Analysis of Heat Flow and Calculation of Current Ratings

The possibility of calculating ratings for uniform bundles on the basis of an analysis of the heat flow in the bundle and to its surroundings was suggested by the physical symmetry of tight bundles of 7, 19, and 37 cables and by the small difference observed between the minimum ratings of tight and loose bundles.

In order to describe the heat flow analytically, it must be possible to assign a

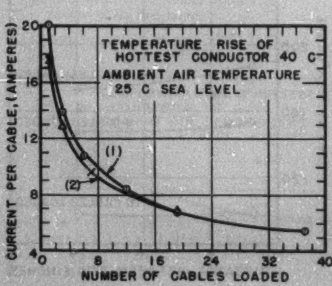


Figure 8 (left). (1) Lowest current ratings for a tight bundle of 37 AN-18 cables. (2) Current ratings for fully loaded, tight bundles of AN-18 cables

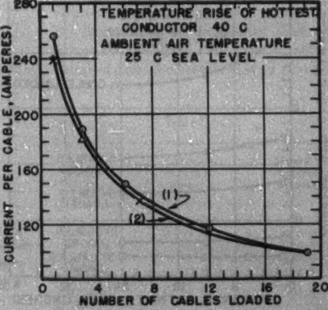


Figure 10 (right). (1) Lowest current ratings for a tight bundle of 19 AN-1/0 cables. (2) Current ratings for fully loaded, tight bundles of AN-1/0 cables

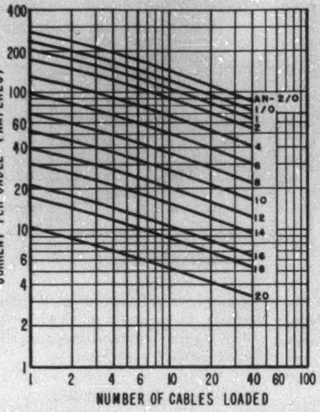
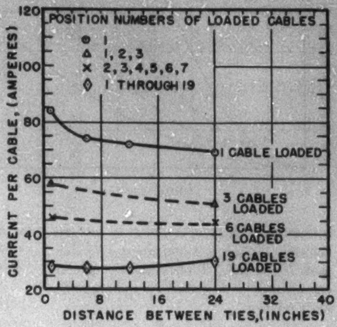


Figure 11 (left). Current ratings for tight uniform bundles of typical aircraft cables when all cables carry the same current. Temperature rise of hottest conductor, 40 degrees centigrade; ambient air temperature, 25 degrees centigrade; sea level

Figure 13 (right). Effect of bundle tightness on current rating for 19 AN-8 cables. Temperature rise of hottest conductor, 40 degrees centigrade; ambient air temperature, 25 degrees centigrade; sea level



definite thermal conductivity, radiation coefficient, and convection coefficient to a bundle. The thermal conductivity K was found by treating the bundle as a cylinder and measuring the quantities required by the heat conduction relation

$$K=\frac{I^2R_1\log_e(b/a)}{2\pi(T_1-T_s)}$$

when current was flowing through the center cable (position 1) only. The symbols in the equation are:

K = thermal conductivity
I = current
R_1 = conductor resistance per unit length at temperature T_1
b = average over-all radius of the bundle
a = radius of the center conductor
T_1 = temperature of the center conductor
T_s = average temperature of the outer bundle surface

The values of K calculated from the data are 0.0040, 0.0050, and 0.0070 watt inch^{-1} centigrade^{-1} for bundles of *AN-18, 8*, and *1/0* respectively.

The radiation coefficient r defined as the heat radiated in unit time by unit area for one degree difference in temperature between the bundle and the surrounding walls, may be computed from the relations governing the net radiation transfer between a body and its surroundings. Because the outer surface of the bundle is indented, there is some uncertainty as to the precise bundle surface area. It can be shown that, as a good approximation, the radiating area per unit length is numerically equal to the circumference of the circle circumscribing the bundle. A plot of the radiation coefficient which assumes a surface emissivity of 0.9 is given in Figure 18. The values for r shown in the figure are obtained from the relation

$$r=\frac{Se(T_{sk}^4-T_{ak}^4)}{T_{sk}-T_{ak}}$$

where

S = Stefan's constant
e = emissivity of cable surface
T_{sk} = average temperature of the bundle surface in degrees absolute
T_{ak} = temperature of the surrounding walls in degrees absolute. It will be assumed that the wall temperature is the same as that of the air at a point where the gradient has become negligible

The convection coefficient, defined as the heat transferred by natural convection in unit time by unit area for one degree difference between the bundle surface temperature and the ambient air temperature, was determined empirically from bundle measurements. The results are

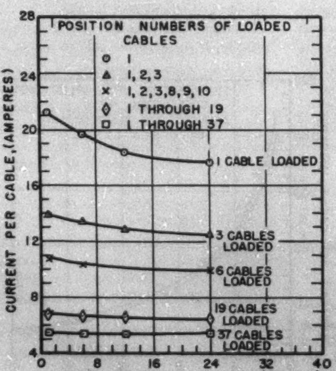
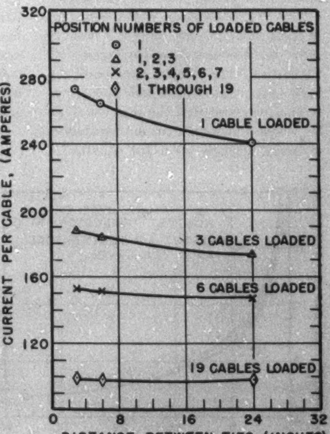


Figure 12 (left). Effect of bundle tightness on current rating for 37 AN-18 cables. Temperature rise of hottest conductor, 40 degrees centigrade; ambient air temperature, 25 degrees centigrade; sea level

Figure 14 (right). Effect of bundle tightness on current rating for 19 AN-1/0 cables. Temperature rise of hottest conductor, 40 degrees centigrade; ambient air temperature, 25 degrees centigrade; sea level



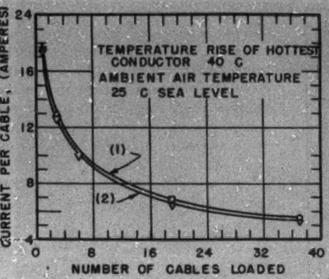


Figure 15 (left). (1) Current ratings for tight, fully loaded bundles of AN-18 cables. (2) Lowest current ratings for loose bundles of 37 AN-18 cables (6-, 12-, or 24-inch ties)

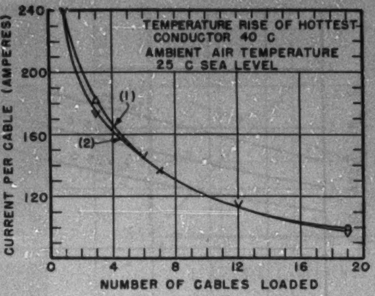


Figure 17 (right). (1) Current ratings for tight, fully loaded bundles of AN-1/0 cables. (2) Lowest current ratings for loose bundles of 19 AN-1/0 cables (6-, 12-, or 24-inch ties)

shown in Figure 19 where the convection coefficient is plotted as a function of the average diameter of the bundle surface.

It is now possible to calculate the current per cable of tight, fully loaded bundles of 7, 19, or 37 cables with the aid of the measured thermal parameters by treating the bundle as a series of concentric cylindrical layers, each of which contributes to the heat flow. The method of calculating the current per cable for a given maximum conductor temperature rise will be described by considering a uniform bundle of 19 cables. A cross section of this bundle, showing the approximate boundaries for heat transfer, is given in Figure 20.

The heat produced by the center cable is conducted to the center of the first layer. The rate of heat flow per unit length in this region is given by the relation

$$I^2R_1=\frac{2\pi K(T_1-T_2)}{\log_e(r_2/r_1)} \tag{1}$$

where

I = current per cable
R_1 = conductor resistance per unit length at temperature T_1
K = effective bundle thermal conductivity
T_1 = temperature of the center conductor, arbitrarily assigned
T_2 = average temperature of the six conductors in the first layer
r_1 = conductor radius
r_2 = radial distance to the center of the conductors of the first layer

The approximation is now made that the heat produced by the six cables of the first layer is added to that from the center cable at r_2. This energy is conducted to the mean center of the second layer r_3. The relation governing the flow of heat from r_2 to r_3 is

$$I^2(R_1+6R_2)=\frac{2\pi K(T_2-T_3)}{\log_e(r_3/r_2)} \tag{2}$$

where

R_2 = conductor resistance per unit length at average temperature T_2
T_3 = average conductor temperature of the 12 cables in the second layer
r_3 = mean radial distance to center of the conductors in the second layer

The further approximation is made that the heat produced by the twelve cables of the second layer is added to that of seven interior cables at r_3, and the total energy is conducted to the outer surface. In the region from r_3 to r_4 we have

$$I^2(R_1+6R_2+12R_3)=\frac{2\pi K(T_3-T_s)}{\log_e(r_4/r_3)} \tag{3}$$

where

R_3 = conductor resistance per unit length at average temperature T_3
T_s = average bundle surface temperature
r_4 = average radius of bundle periphery. An approximate value of r_4 may be found by subdividing a representative section of the periphery into a number of equal segments and taking the average of the radial distances to the ends of each segment. For example, in the case of the bundle of 19 cables shown in Figure 20, *ABC* and *CDE* are examples of representative sections. If *ABC* is subdivided into a number of equal elements, the average of the distances from the bundle center to the ends of each segment (two distances for each segment) gives the average radius of the bundle periphery. The same result is obtained if the computation is made for *CDE* or any other of the representative sections

At the outer bundle surface, heat is transferred by convection and radiation to the surrounding air and walls. The rate of heat transfer can be stated as

$$I^2(R_1+6R_2+12R_3)=2\pi(r_4h+r_5r)(T_s-T_a) \tag{4}$$

where

h = convection coefficient
r_5 = radius of the circle which circumscribes the bundle $=2.5r_2$
r = radiation coefficient
T_a = ambient air temperature, that is, temperature of the air at a point where the temperature gradient has become negligibly small

The current required must satisfy equations 1, 2, 3, and 4 simultaneously. In these equations the unknown quantities, in addition to the current, are T_2, T_3, T_s, R_2, R_3, h, and r, the latter four quantities being dependent on the three preceding. In order to find I, the method of successive approximations is used, that is, one value of I is assumed and substituted in turn in equations 1, 2, and 3. By this means one determines a set of values of T_2, T_3, T_s, R_2, and R_3. The values of r and h corresponding to T_s-T_a

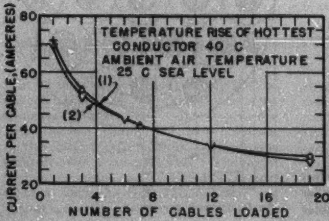


Figure 16. (1) Current ratings for tight, fully loaded bundles of AN-8 cables. (2) Lowest current ratings for loose bundles of 19 AN-8 cables (6-, 12-, or 24-inch ties)

Table IV

Bundle Designation	Calculated Current, Amperes	Measured Current, Amperes
7-18-1	9.90	9.65
19-18-1	6.90	6.90
37-18-1	5.45	5.50
7-8-1	41.5	40.6
19-8-1	28.4	29.5
7-1/0-1	135	137
19-1/0-3	94.5	99.0

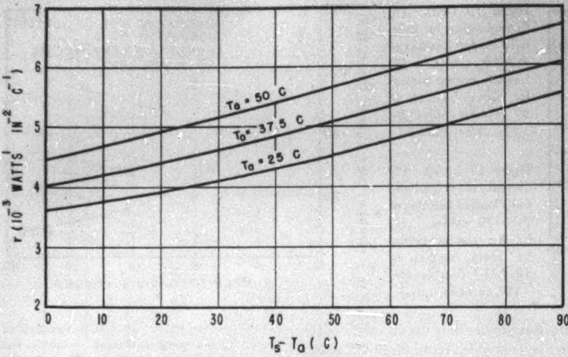


Figure 18. Radiation coefficient r versus average surface temperature rise $T_s - T_a$ for three ambient temperatures T_a

and bundle diameter now may be selected from Figures 18 and 19. Finally, a value of I can be computed with equation 4. If this value agrees with the one assumed initially, the correct current has been found. If the two currents differ, other assumed values must be tried and the calculations repeated. The number of trial values required may be limited to two or three by plotting the current calculated by equation 4 as a function of the assumed current. The intersection of this curve with the straight line, I (calculated) $= I$ (assumed), gives the desired current. The extension of the method to bundles consisting of an integral number of layers is self-evident. Computation details for a bundle of 19 *AN-18* cables are given in Appendix II.

The current necessary to produce a 40-degree-centigrade rise in the hottest conductor was calculated for a series of bundles by the method described. The results of these calculations and the currents observed experimentally are given in Table IV.

Conclusions

Detailed and quantitative information has been made available concerning several of the factors affecting current rating and heat transfer. Some of the data, for example, bundle convection coefficients and the range of thermal conductivity, are applicable not only to uniform bundles but also to mixed bundles (bundles composed of cables of more than one size). Two factors which cannot be expressed easily in quantitative form, the position of the loaded cables and bundle tightness, can be taken into account for uniform-bundle ratings. The conclusions which follow are valid for the range of cable and bundle size investigated. A given rise in the temperature of the hottest conductor is to be assumed.

1. The current per cable decreases as the number of cables loaded increases and obeys the approximate relation .

$$I_n = \frac{I_1}{n^{\lambda}} \qquad (0.29 < \lambda < 0.33)$$

for tight, fully loaded, uniform bundles.

2. The current per cable for a given number loaded depends on the position of the loaded cables within the buudle. Variations up to 30 per cent in current per cable for a given number loaded can be produced by changing the positions of the loaded cables within the bundle.

3. The current per cable for a given number loaded increases as the total number of cables in the bundle increases and approaches a maximum value.

4. For the tight-bundle condition, the current per cable for a given number of cables loaded is a minimum when the number of cables loaded equals the total number of cables in the bundle.

5. The minimum current per cable for each number of cables loaded for the case of the large, loosely bound bundle (19 or 37 cables) was as much as six per cent below the minimum tight-bundle values.

6. The effects of the position of the loaded cables within the bundle and of the total number of cables in the bundle on current rating for a given number loaded may be neglected if ratings are based on conditions which give the minimum current values. The condition for tight bundles which gives the poorest rating for a given number loaded is the fully loaded bundle. Although the loose-bundle condition gives lower current

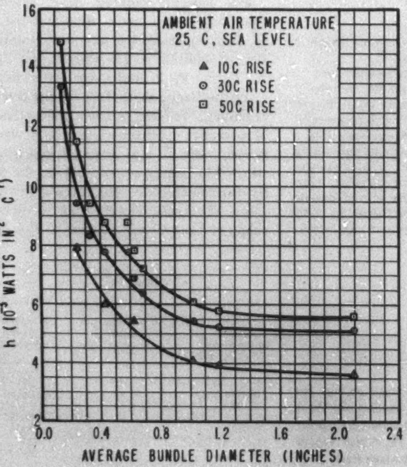


Figure 19 (left). Convection coefficient h versus average bundle diameter for three values of average surface temperature rise

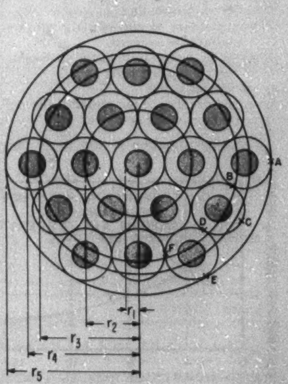
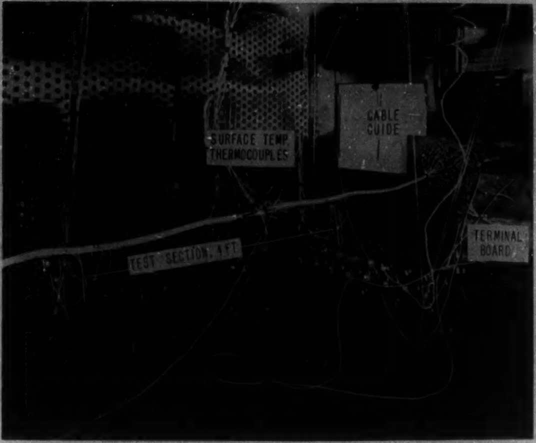


Figure 20 (right). Heat transfer boundaries for a bundle of 19 cables



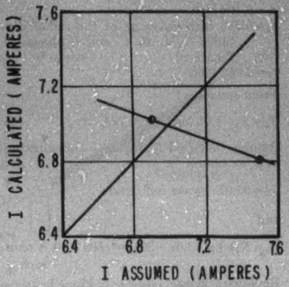


Figure 21 (left). Experimental setup for measurements on bundles of aircraft cables

Figure 22 (above). Graphical solution for current per cable for bundle 19-18-1

values than the tight-bundle condition, ratings for uniform bundles may be based on those for the tightly laced condition provided that a valid derating factor for the loose-bundle condition is known. The magnitude of derating for the range of loose bundles investigated has been indicated.

7. A method for calculating the current per cable necessary to produce a given temperature rise for uniform bundles of 7, 19, or 37 cables has been described. Current ratings for bundles with a total number of cables between 1 and 37 may be found by interpolation from the calculated values. An estimated accuracy of 10 per cent is attainable for cylindrical-shaped, uniform bundles containing up to 37 cables.

8. Uniform bundle ratings may be applied directly to mixed bundles in which the cables loaded simultaneously are of a single size.

9. Temperature rise for fully loaded, tight, uniform bundles is directly proportional to the square of the current per cable for the range of temperature rise observed.

10. Although certain parameters have been fixed throughout the discussion (for example, the ambient air temperature at 25 degrees centigrade, the maximum conductor temperature rise at 40 degrees centigrade, and the air pressure at sea-level value) the results and conclusions presented here may be extended easily to other values.

Appendix I. Measurements

In all tests, measurements were taken of current, conductor resistance, conductor temperature, and ambient air temperature after thermal equilibrium had occurred. In the case of tight bundles, surface temperature was measured also. The equilibrium condition was arbitrarily defined to exist when the temperature rise of the hottest conductor changed less than 0.1 degree centigrade in 15 minutes, provided that both the current and the ambient air temperature remained constant. For the range of bundles investigated, the time required for equilibrium varied from 45 minutes to 4 hours.

The source of current, a 6,000-ampere-hour battery, gave a current stable to within 0.04 per cent once thermal equilibrium of the circuit was reached.

All tests were conducted in a draft-free, thermally insulated chamber measuring 9 by 12 by 8 feet. The chamber was able to maintain a nearly constant ambient temperature in spite of the fact that the power dissipated by the bundles approached 1,000 watts. The maximum rate of change in ambient temperature in the chamber was observed to be less than 0.2 degrees centigrade in 15 minutes.

The bundles were suspended as shown in Figure 21 so that the relative positions of the component cables were maintained throughout the bundle length. To assist in assembling the bundle, cable guides were used. Since all the loaded cables were to carry the same current, the cables were brought to terminal boards, which provided a convenient means of connecting any combination in series.

The test section was situated at the center of the 10-foot bundle. It is believed that the 8 feet of bundle at each end of the test section was sufficient to insure a constant temperature along the length of each cable in the region of the test section.

Cable resistances were determined by comparing the measured voltage drops across the 4-foot test sections with that across a Leeds and Northrup standard shunt. Conductor temperature was then calculated by the change-in-resistance method. Tests on sample cables indicated that the temperature coefficient of resistance for 20 degrees centigrade was 0.00395. It is estimated that measurements of conductor temperature rise were accurate to within 1 per cent.

It was not practical to measure the temperature of every conductor in the larger bundles. In most cases, temperatures of the center conductor and three conductors in each layer were measured. For example, in bundle 19-8-1 temperatures of cables 1, 2, 3, 5, 8, 9, and 14 were measured.

Considerable difficulty was experienced in devising a method of measuring surface temperature. Since temperature varied over the bundle circumference some means of averaging these temperatures was necessary.

In addition, the temperature gradient of the air close to the bundle surface was very large, so that any temperature measuring device had to be an integral part of the surface. After several different techniques had been tried it was decided to employ copper-constantan thermocouples constructed from American Wire Gauge 40 wire. These were imbedded just beneath the surface at the center of the exposed portion of each outer cable. It is estimated that the surface temperature measurements were accurate to within 3 degrees centigrade.

The voltages from the temperature measuring devices and from the standard shunts were measured by a Leeds and Northrup type *K-2* potentiometer with a Leeds and Northrup number 2285 galvanometer to indicate balance. Conductor and over-all diameters were determined by averaging 20 micrometer readings.

Appendix II. Calculation Details for Bundle 19-18-1

To determine: *I*, current per cable, when every cable in the bundle carries current simultaneously.

Given:

$T_1 = 65$ degrees centigrade
$T_a = 25$ degrees centigrade
Conductor resistance per inch at 20 degrees centigrade $= R = 0.503 \times 10^{-3}$ ohm inch^{-1}
Temperature coefficient of resistance $= \beta =$ 0.00395 at 20 degrees centigrade
$K = 0.0040$ watt inch^{-1} centigrade^{-1}
$r_1 = 0.0242$ inch
$r_2 = 0.104$ inch

$r_3=0.194$ inch
$r_4=0.216$ inch = average of the radial distance to A, B, C, and D shown in Figure 20
$r_5=0.259$ inch

From equation 1

$$T_2=T_1-I^2R_1\log_e(r_2/r_1)/2\pi K \tag{5}$$

Assume that $I=7.5$ amperes. Substituting in 5, we obtain

$T_2=63.07$ degrees centigrade

and

$R_2=R[1+\beta(63.07-20)]=0.589\times10^{-3}$ ohm inch^{-1}

Equation 2 gives

$$T_3=T_2-\frac{I^2(R_1+6R_2)}{2\pi K}\log_e(r_3/r_2) \tag{6}$$

$T_3=57.31$ degrees centigrade

and

$R_3=0.577\times10^{-3}$ ohm inch^{-1}

Equation 3 gives

$$T_s=T_3-\frac{I^2(R_1+6R_2+12R_3)\log_e(r_4/r_3)}{2\pi K} \tag{7}$$

$T_s=54.57$ degrees centigrade

I is now calculated with equation 4.

$$I=\left[\frac{2\pi(r_4h+r_5r)(T_s-T_a)}{R_1+6R_2+12R_3}\right]^{1/2} \tag{8}$$

which is evaluated for

$(T_s-T_a)=54.57-25.0=29.57$ degrees centigrade

The radiation and convection coefficients now can be obtained from Figures 18 and 19.

r, for $T_s-T_a=29.57$, $=0.00413$ watt inch^{-2} degrees centigrade^{-1}
h, for $T_s-T_a=29.57$ and bundle diameter $=2r_4$, $=0.0078$ watt inch^{-2} degrees centigrade^{-1}

Substituting in equation 8 we find

$I=6.81$ amperes

Since the value for I assumed and that calculated by equation 8 differ, additional trials are required. It is found that when I is assumed to be 6.9 amperes, the current calculated by equation 8 is 7.02 amperes. If the calculated values are plotted as a function of the assumed values, and if the straight line I (calculated) $=I$ (assumed) is drawn on the same graph, the intersection at 6.99 amperes gives the desired current. This result is shown in Figure 22.

No Discussion

Electric Glass Welding

M. R. SHAW, JR.
MEMBER AIEE

THE ELECTRIC welding of metal is a relatively new and rapidly growing process in the metal fabrication industries. Even newer and currently enjoying a more rapid expansion is the electric welding of glass. While there is some similarity between the techniques employed, the widely differing physical properties of the two materials make the electrical requirements quite distinct.

Type of Electric Heat Used

It may be said that all forms of electric heat[1-3] have been used in welding glass. Probably the most expensive type of electric energy is that in the ultrahigh-frequency portion of the spectrum. Electric energy at this frequency has the ability to produce appreciable heat in most dielectrics and has been so used on glass under certain circumstances. The harder glasses maintain their dielectric characteristics up to moderately high temperatures. It is in this region that dielectric loss methods of heating may be used as part of an electric welding process. Radio frequency energy also may be used for magnetic induction heating of materials having fair conductivity. Since glass at very high temperature has moderate conductivity, it may be heated for welding by the high-frequency induction method. Generally, however, unless the dielectric loss or induction heating method is uniquely suited for the application, other types of electric heating will be more practical. The kind of electric heating found most useful for glass welding on a production basis is the simple resistance loss accompanying the passage of current through the glass between opposed electrodes. The source of energy for this type of resistance heating may be the 60-cycle power line or, under certain circumstances a high-frequency generator. Undoubtedly time will bring the development of energy sources for dielectric and induction heating which will greatly change their economics. For the present, good use is being made of the simpler method, and that will be the subject of the remainder of the paper.

Resistance Heating in Glass

Since resistance heating is to be examined closely, it is appropriate that the electrical properties of glass be known. In Figure 1 is shown a resistivity/temperature curve for a borosilicate glass. This graph covers the interval of room temperature to 1,500 degrees centigrade which is approximately the same as that involved in welding this particular glass. The extreme range of resistivity is immediately evident. The resistance of a given path will decrease by a factor of 10^{14} as the temperature increases during the welding cycle from the room-temperature value to the fusion point. What this means in terms of amperes, volts, and power is indicated in Figure 2, where glass and metal are compared. Here the voltage and current necessary to deliver 100 watts to a particular load are tabulated for different temperatures. At 20 degrees centigrade, in the case of steel, the current and voltage requirements are easily within the realm of practicability. This is not so, however, for the glass. In this instance a voltage is required which is many times that which the normal surrounding atmosphere would support. The cube would be heated by the resulting flashover. Arc heating of glass has been employed but at less than a third of a billion volts.

The next two lines in the tabulation show the conditions prevailing at successively higher temperatures. A slight upward adjustment in voltage is necessary for the steel to compensate for its increase in resistivity. Going to the 500 degree-centigrade temperature in the case of the glass changes the situation from one that is impossible to one that is reasonable. At this temperature a borderline condition obtains. With well-designed electrodes it probably would be possible to restrict the current path to the glass. Accom-

---

Paper 52-320, recommended by the AIEE Electric Welding Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Middle Eastern District Meeting, Toledo, Ohio, October 28–30, 1952. Manuscript submitted November 28, 1950; made available for printing September 16, 1952.

M. R. SHAW, JR., is with the Corning Glass Works, Corning, N. Y.

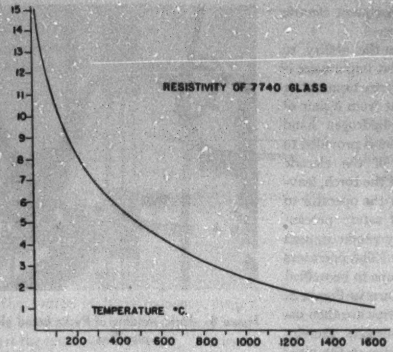


Figure 1 (left). Resistivity / temperature characteristic of a borosilicate glass

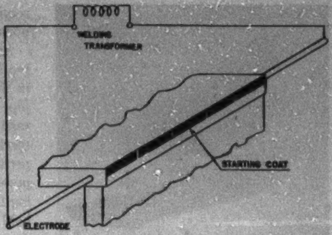


Figure 3 (right). Application of starting coat on glass to be welded

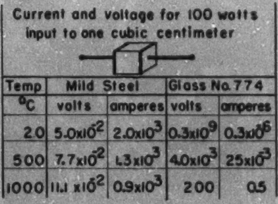
plishing this restriction, while not impossible, is still hardly practical. Going on now to 1,000 degrees centigrade it is observed that a potential of only 200 volts is required. At this point there is no question about being able to confine the current to the glass. In the region between 500 degrees centigrade and 1,000 degrees, therefore, it is possible to begin to employ electric resistance heating on a practical basis in this particular material.

There is no special significance in the value of 100 watts per cubic centimeter used in this illustration. Due to the negative resistance/temperature coefficient of glass, the electric heating process becomes self-accelerating when the power input exceeds the heat loss. This circumstance has both good and bad aspects. The favorable fact is that the direction of change is such as to promote the passage of current through the load rather than through the surrounding atmosphere. The disadvantage is that a run-away condition is quickly reached and provision must be made to take care of it. In the preceding sample, the electric heating process would become self-sustaining at any combination of temperature and voltage where the power delivered was equal to the heat losses. This balance could be maintained or changed in either direction by proper adjustment of the circuit elements. Experience has shown that the integrated effect of all factors, including practical limits on the voltage level and changing dielectric strength of the surrounding atmosphere, is to establish a minimum temperature of about 500 degrees centigrade for starting the electric heating cycle in the harder glasses, which are of most interest. The problems associated with electric glass welding are now fairly well defined. Some preheating device is necessary for bringing the glass temperature into the starting range, and a method of controlling the power is required.

Starting Coat for Preheat

Several ways of preheating have been developed but there are two which are used the most. One employs colloidal graphite painted on the glass to form a semiconducting stripe.[4,5] It may be painted on one of the pieces to be joined, as shown in Figure 3, or on a third piece of glass which serves as a filler rod, as shown in Figure 4. It will be noticed that in both cases there are small gaps equally spaced along the coat. These gaps play an important role in the starting mechanism. When power is applied at the electrodes at the proper voltage level, a series of small arcs forms along the rod at the gaps in the coat. These arcs, when properly controlled, hug the glass welding rod very closely and as a result impart to it considerable heat. The effect of this is to raise the glass temperature with resulting decrease in resistance in the zone under the arcs. Concurrently, the coating on the glass burns back from each end of the arc. These two actions, by decreasing the resistivity of the glass and stretching out the arc path, work together to bring about the transfer of the current from the surrounding atmosphere to the welding rod. Once the current is well established in the glass, the power is allowed to increase and the fusion takes place. Because of the negative resistance/temperature coefficient of the material, the tendency of the current path is to increase in cross section relatively slowly. By a rapid acceleration of the power input, the filler rod can be liquified and welded to the pieces being joined before the latter are appreciably distorted. This high degree of heat localization is often an important advantage of the electric process.

The beginning and end of a 15-inch filler rod weld are shown in Figures 5 and 6. The electrodes contact the ends of the rod until immediately before the power is turned off. At this point they are swung back, drawing an arc which melts out the mark which otherwise would be left at the contact spot. In Figure 7 is shown a group of typical sections built up by this method. The glass used in making these was 1/4 inch thick.



Current and voltage for 100 watts input to one cubic centimeter

Temp	Mild Steel		Glass No. 774	
°C	volts	amperes	volts	amperes
20	5.0×10^{-2}	2.0×10^{3}	0.3×10^{9}	0.3×10^{-6}
500	7.7×10^{-2}	1.3×10^{3}	4.0×10^{3}	25×10^{-3}
1000	11.1×10^{-2}	0.9×10^{3}	200	0.5

Figure 2 (left). Comparison of voltage and current requirements for heating metal and glass

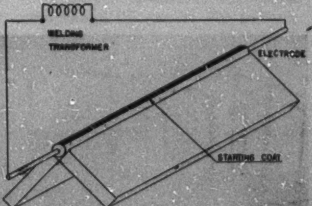


Figure 4 (right). Application of starting coat to glass filler rod

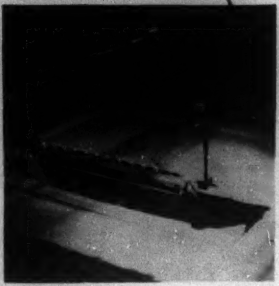


Figure 5. Beginning of 15-inch filler rod weld

Flame Preheat for Starting

When the welds to be made are cylindrical in shape, a flame is often used as a preheating medium.[6] The heating of the flame is principally on the surface of the glass. Increasing the temperature of the flame to increase the heating rate serves only to promote damage of the glass surface by volatilization. However, flames of moderate intensity may be used to initiate the electric heating process by bringing the glass temperature into the 500 degrees centigrade zone. The electric heating that follows is then volume heating thereby permitting a much faster heating rate. Once again a high degree of localization is realized and the zone initially subjected to the preheat flames receives most of the subsequent electric power.

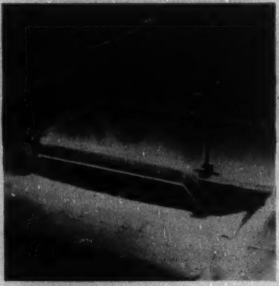


Figure 6 (upper left). End of 15-inch filler rod weld

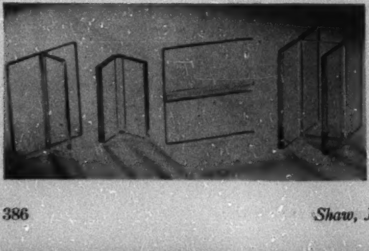


Figure 7 (left). Typical sections made by filler rod method

An application where the ability to localize the heat is of great importance is in the field welding of Pyrex brand glass pipe line. Flame preheat from a pair of easily manipulated oxy-hydrogen hand torches is used. Insulation is provided to permit the connection of the electric power at the flame end of the torch, leaving the other end safe for the operator to hold. As an additional safety precaution, only high-frequency power is used for this type of welding. The operators position the ends of the pipe to be welded by means of a jig, as shown in Figure 8. The two flame electrode tips are then oscillated through 180-degree arcs, thereby heating the entire perimeter of the pipe. The flames serve a second purpose by forming a flexible brush to lead the electric power to the load. This is a desirable feature in the hand-manipulated process because it is not easy to maintain the electrode spacing with any accuracy. When the hot stripe is developed, the power is turned on and the ends of the pipe brought up to the fusion temperature as they are worked together. A band approximately 1/4 inch wide is all that is actually hot enough to flow. This band is supported by surface tension forces and slightly increased internal pressure while the union takes place. As a result, it is easy for an unpracticed person to maintain complete control of the glass and turn out welds beyond the capacity of even the skilled glass artisan working with flames alone. This was demonstrated in the construction of the atomic plant at Oak Ridge, Tenn., where there were at one time 25 men recruited from the ranks of the pipe fitters engaged in the installation of $7^{1}/_{2}$ miles of glass pipe line using the Corning electric glass welding system.[7]

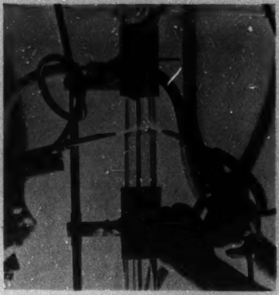


Figure 8. Field welding of Pyrex brand glass pipe line

With increasing diameter it becomes advantageous to place the parts to be welded on a lathe. The work may then be rotated in the preheat flames to assure uniformity of heating. Because of the greater size of the parts handled on the lathe, the operators usually use a battery of gas-oxygen flames for this purpose instead of depending on the heat from the electrode flames. Since the large work done on the lathe places greater demand on the power source, the 60-cycle supply generally is used because of its economy and greater reliability. The lathe set up for welding a large television bulb is shown in Figure 9. A substantial proportion of the television bulbs reaching the market today are made by the electric welding process.

Power Supply and Control

SIXTY-CYCLE WELDING

Sixty-cycle potentials as high as 20,000 volts have been required to handle some of the large jobs encountered. A distribution-type transformer is usually satisfactory as a source of high voltage and has the advantages associated with standardized components. A certain amount of circuit reactance is required to stabilize arcs in either the starting-coat method or the flame-electrode method. The requirements for inductive reactance for this purpose and also for providing power control are adequately satisfied by saturable core reactors in most cases. A simplified circuit diagram is given in Figure 10.

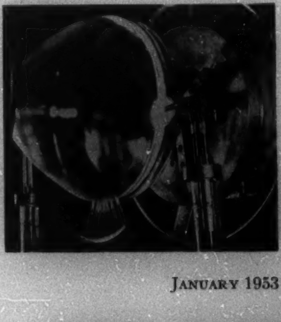


Figure 9 (right). Electric welding of large glass television bulb

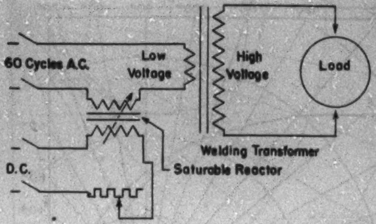
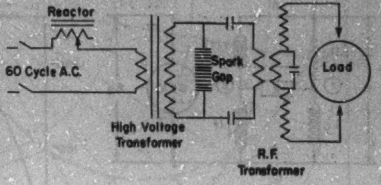


Figure 10 (left). Basic circuit for 60-cycle glass welding

Figure 11 (right). Basic circuit for high - frequency glass welding



HIGH-FREQUENCY WELDING

The source of high-frequency energy most generally used for this type of welding is the spark-gap converter. The usual induction heating type of equipment is not suitable without modification because a high voltage rather than a high current output is desired. A basic circuit drawing for a suitable converter is shown in Figure 11. The frequency used is not important, something in the range of 400 kc being usual. Power control is generally an on-off arrangement but load matching circuits have been designed for delivering maximum power during the important part of the heating cycle.

Summary

The extension of the electric welding process to include glass has opened up a vast new field to the electrical art. Hundreds of kilovolt-amperes of installed capacity are now turning out highly technical products at a rate such as to permit them to reach the mass market. Electric glass-welding processes have been developed for handling all kinds of joints in both straight and curved bodies. The advantages of electric glass welding may be attributed to the following:

1. It is possible to develop higher temperature within the body than with flames alone.
2. A faster heating rate is possible without danger of destroying the surface of the material.
3. The highly localized nature of the heat eliminates the possibility of distortion in the main body of the parts.
4. Accurate control of the heat is possible, resulting in better quality of finished product.
5. The nature of the process is such that it may be carried out by relatively unskilled personnel.

References

1. HIGH FREQUENCY ELECTRIC GLASS WELDING, E. M. Guyer. *Transactions*, Electrochemical Society (New York, N. Y.), volume 79, 1941, pages 187–98.

2. ELECTRONIC WELDING OF GLASS, E. M. Guyer. *Electronics* (New York, N. Y.), volume 18, June 1945, pages 92–96.

3. ELECTRIC WELDING OF GLASS, E. M. Guyer. *Electrical Engineering*, volume 67, May 1948, pages 446–47.

4. DOUBLE GLAZED WINDOW AND METHOD OF MANUFACTURE, Edwin M. Guyer, Jesse T. Littleton, Morton R. Shaw, Jr. United States Patent *Number 2,389,360*, November 20, 1945.

5. METHOD OF WELDING GLASS, Morton R. Shaw, Jr. United States Patent *Number 2,470,376*, May 17, 1949.

6. GLASS HEATING AND WORKING, Edwin M. Guyer. United States Patent *Number 2,306,054*, December 22, 1942.

7. MILES OF GLASS PIPING FABRICATED IN FIELD AT OAK RIDGE, B. W. Whitehurst. *Chemical and Metallurgical Engineering* (New York, N. Y.), volume 53, July 1946, pages 112–15.

No Discussion

Stabilizing Elements in Aircraft D-C Systems

D. G. SCORGIE
ASSOCIATE MEMBER AIEE

D. H. SCHAEFER
ASSOCIATE MEMBER AIEE

TWO RECENT studies of aircraft generator regulator systems have come from this laboratory. One[1] analyzed the carbon-pile regulator and determined its transfer function; the other[2] derived a generator-transfer function and determined stability criteria for a regulator-generator combination. These studies were concerned primarily with design parameters of the regulator and generator but did not answer the question of how to improve matters once the regulator and generator design were fixed. In particular, an understanding of the function and limitations of the standard feedback resistor was desired. Therefore, a mathematical and experimental investigation was undertaken to obtain the transfer functions of systems with feedback. The mathematical determination of these functions was straightforward. They were, however, too cumbersome and complicated in their exact form to be of great help in noting the effect of changes of parameters, primarily because there are no buffer amplifiers allowable between components. In some cases, but not all, reasonable approximations have been derived. There are, however, mathematical generalizations in the frequency plane one can fall back upon for the purpose of understanding or designing stabilizing networks. These generalizations, in addition to transfer-function measurements, form the basis of work which has been done here.

In order to measure the loop-transfer functions of the systems with feedback,

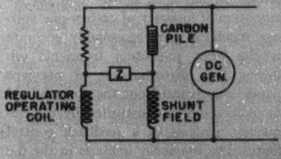


Figure 1. Generator system under study

Paper 52-330, recommended by the AIEE Air Transportation Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Middle Eastern District Meeting, Toledo, Ohio, October 28–30, 1952. Manuscript submitted June 17, 1952; made available for printing September 5, 1952.

D. G. SCORGIE and D. H. SCHAEFER are both with the Naval Research Laboratory, Washington, D. C.

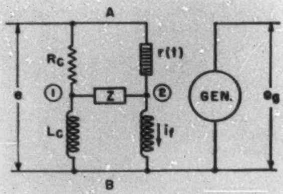


Figure 2. Open loop with respect to generator

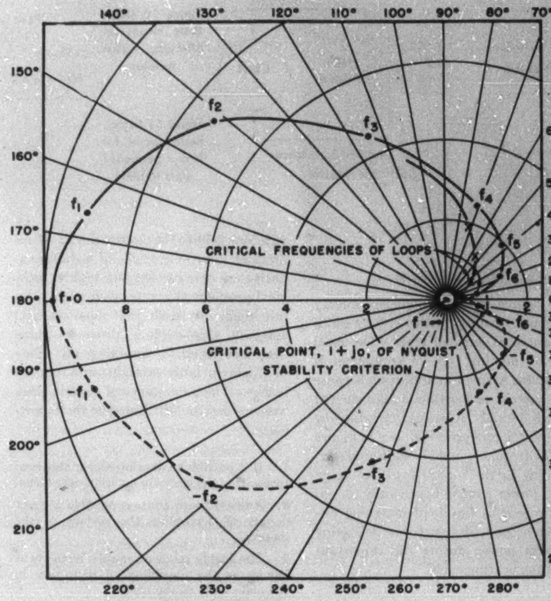


Figure 3. Typical open loop-transfer function

a new measuring circuit had to be developed, see Appendix I. With this electronic circuit it has been possible to obtain transfer functions of both stable and unstable generator-regulator combinations at various regulator adjustments and with various feedback components. Data were taken simulating the no-load full-speed condition, which is known empirically to be the condition of poorest stability.

It is known that the improvement of present-day systems by a feedback resistor is limited in its effectiveness. Indeed, if the resistor value is too low, both d-c regulation and system stability are ultimately destroyed. One alternate approach which has been used is to restrict the stabilizing effect with resonant circuits to a fairly narrow frequency band, leaving the steady-state regulation undamaged. In the course of the investigation, it was found that eddy currents in the generator affect the stability of a system in an important manner. Measurements of a typical eddy-current effect, made in the frequency plane, are presented here. The eddy-current effect is revealed by the lagging of generated voltage behind field current, and also by the fact that the input impedance of the field circuit is not that of a resistance-inductance circuit.

General Criteria for System Stabilization

Figure 1 shows the closed-loop feedback system to be studied. It consists of a d-c generator controlled by a carbon-pile voltage regulator and contains a feedback impedance *Z* which is usually a resistor. It has been shown[2] that the assumption of linearity of carbon-pile voltage regulator gives reasonably accurate results when predicting transients following load disturbances. A previous paper[2] also pointed out the powerful aid to understanding constituted by the frequency analysis of a feedback control loop. The frequency analysis in the d-c system is accomplished by considering only a small variation superimposed upon the large d-c initial conditions. In the earlier paper no auxiliary feedback stabilizing elements were considered. Because of the feedback elements, the closed loop in this paper was opened at a different point than in the previous paper[2], and the analysis of the system carried out in a somewhat different manner.

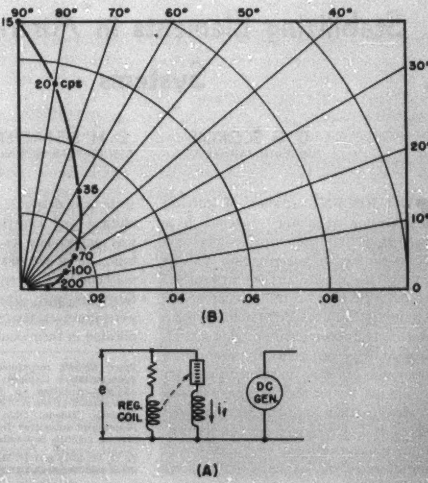
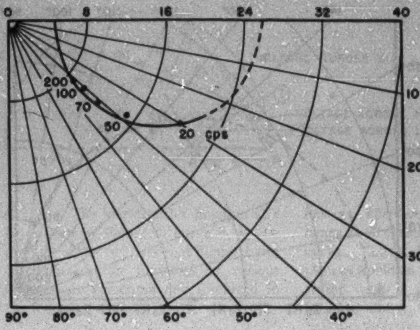
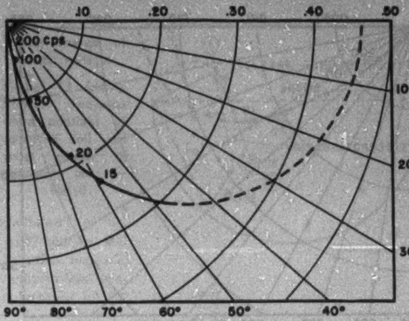


Figure 4. $i_f/e(\omega)$ with regulator operating, no feedback





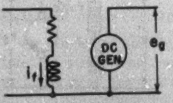
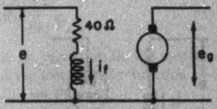


Figure 5 (left). $e_g/i_f(\omega)$ for d-c generator

Figure 7 (right). $e_g/e(\omega)$ with regulator inoperative



For instance, it became necessary to take into account the fact that there are two active elements in the system. The generator is obviously active. When looking at variations from the d-c references, the carbon-pile regulator is active by virtue of the fact that the generator supplies a steady-state background carbon-pile current. The equivalent voltage generator, due to superimposed variations of carbon-pile resistance, is exactly analagous to the equivalent voltage in vacuum-tube plate circuits when grid bias alters the tube impedance. A full discussion of the difference in approach and the reasons for this difference is contained in Appendix II.

In this paper the closed loop has been broken, as shown in Figure 2. The function to be measured is the loop-transfer function, e_g/e, with respect to one active element, the generator. (Throughout the report, lower case letters indicate incremental changes of the variable from initial steady-state conditions.) For the measurements, this function was broken into two parts, i_f/e and e_g/e_f, which were measured separately. The latter part was frequency-dependent solely because of eddy currents in the generator field pole structure. To obtain e_g/e, the two functions, i_f/e and e_g/i_f, were multiplied together.

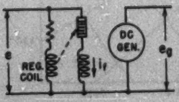
The measurement of i_f/e was not possible with existing equipment, due to the difficulty in providing a stable d-c background upon which to superimpose the small sinusoidal variations. Consequently, new circuitry was devised which will be discussed in Appendix I.

A typical curve of $e_g/e(\omega)$ is shown on the polar plot, Figure 3, with frequency as a parameter. The determination of stability and transient characteristics from this plot is discussed in many places, including a paper by R. L. Mills[1] and a book by H. W. Bode.[4] With two active elements in our system, instability is determined by first deactivating one and examining the stability of the system with the other active. Only the a-c component of voltage must be deactivated. For instance, suppose that instead of the generator in Figure 1 a battery, whose generated voltage will not of course change, is substituted. With no feedback element Z connected, this change also insures that the regulator operating coil current will remain constant and the system in this condition will not oscillate. System instability is determined by reactivating the generator and examining whether or not the plot e_g/e encircles the $1+j0$ point. The plot first must be extended to negative frequencies by forming the mirror image of the positive frequencies about the real axis. If the total curve encircles $1+j0$ point more times in the clockwise direction than in the counter-clockwise direction, the system is unstable, that is, it is characterized by self-oscillations of growing magnitude. If not, the oscillations following a disturbance will die away eventually. Among the systems which are not unstable by this definition, varying degrees of stability are spoken of, although the concept can be defined only roughly. A transient response to a sudden disturbance in the system will, in general, be a damped sinu-

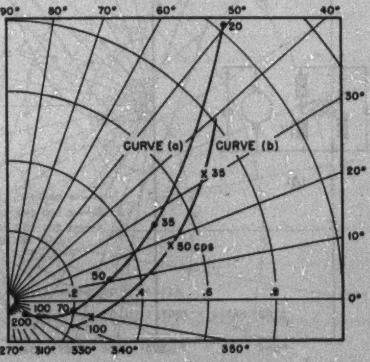
Figure 6. $e_g/e(\omega)$ with regulator operating, no feedback

A. Product of Figures 4 and 5

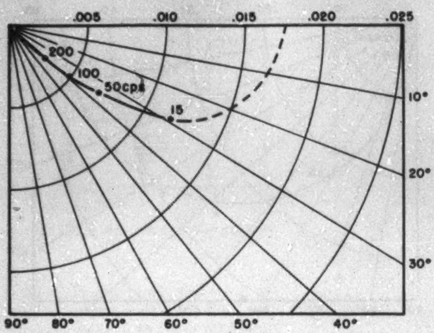
B. Pile screw readjusted for improved steady-state regulation

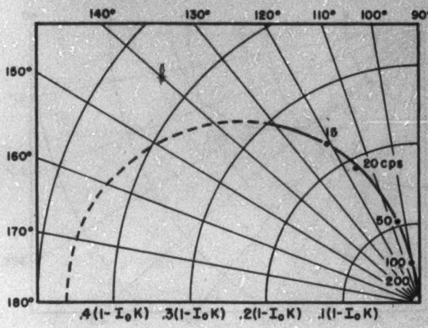


(A)



(B)





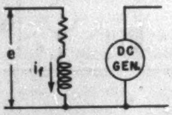
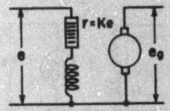


Figure 8 (left). $i_f/e(\omega)$ with regulator inoperative

Figure 9 (right). $e_g/e(\omega)$ where regulator has flat frequency response



soid. The frequency of the damped sinusoid corresponds roughly to the frequency at which the curve of Figure 3 approaches closest to the $1+j0$ point. Given a particular input disturbance, the closer the loop-transfer function approaches to the $1+j0$ point, the greater will be the magnitude and the slower will be the damping out of the response. It is known, for instance, that the system whose loop-transfer function is the crossed curve of Figure 3 recovers from a disturbance more rapidly than that whose full loop-transfer function is shown.

When the feedback impedance Z of Figure 1 is connected and the generator is deactivated, an analysis must be made to determine stability of the minor loop before Nyquist's stability criterion previously discussed is valid. The case of primary interest is when Z is the usual pure resistance. An experimental determination of stability under these conditions is next to impossible. The analysis is very involved but leads to the conclusions that in order to be stable:

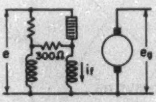
1. The slope of the carbon-pile resistance versus operating-coil voltage must be positive.
2. The impedance Z should be a full order of magnitude greater than the impedance of the generator field winding.

There is no worry about going into the matter in too great detail. The primary concern is with improving stable systems which are known to have poor transient response, or which border upon instability. Therefore, only loop-transfer functions will be considered with respect to the generator.

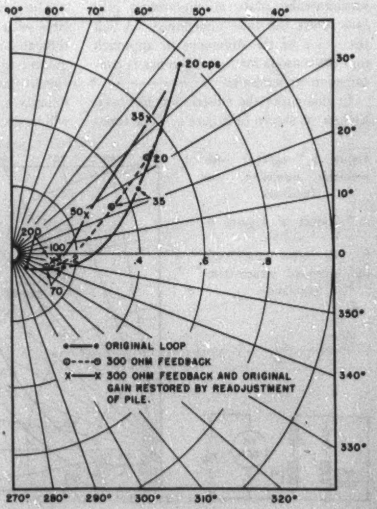
Most transient-input disturbances have frequency components which fall off in magnitude at least as the inverse first power of frequency. Therefore, it is usually best to make the characteristic frequency of the system as high as possible. To improve the transient response of the system, the desired operation is to move the loop plot away from the $1+j0$ point as shown by the crossed curve, Figure 3. It is possible to accomplish this in many ways by the addition of feedback and cascade elements. At the same time, the added element must not affect the steady-state regulation (zero frequency response) adversely.

With the equipment described in Appendix I, the transfer functions previously discussed can be measured. Figure 4 shows a typical measured value of $i_f/e(\omega)$. Figure 5 shows similar measurements of

Figure 10. $e_g/e(\omega)$ with 300-ohm feedback resistor, regulator operating



(A)



(B)

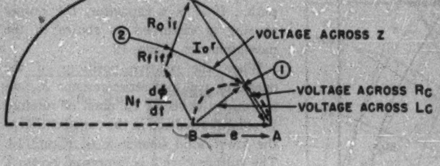
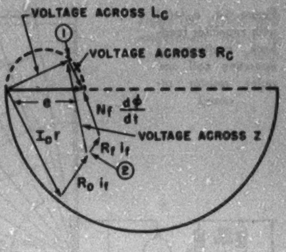


Figure 11 (left). Vector diagram of system voltages

Figure 13 (right). System voltages when positions of operating coil and series resistor of regulator are reversed



$e_g/i_f(\omega)$. The loop-transfer function $e_g/e(\omega)$ shown in Figure 6, curve *a*, is the product of the two curves of Figure 4 and 5. Loop-transfer functions of various systems are presented and discussed in the following section.

Experiments in System Stabilization

RESPONSE OF THE D-C GENERATOR

We can follow step by step the synthesis of a regulating loop and examine each change of characteristic experimentally. The principal building block is, of course, the d-c generator.

The test result of Figure 7 is the transfer function $e_g/e(\omega)$ of a generator alone with no regulator operating. For the range of frequencies measured, it has the form of a simple time lag element. If there were no eddy currents whatsoever in the iron structure, the time constant of this loop could be determined from the resistance and self-inductance of the shunt field, since generated voltage and field current would be in phase. In an actual generator, however, no single constant field inductance can be measured because of the eddy currents. Instead, $i_f/e(\omega)$ plots as shown in Figure 8. From this figure it can be calculated that the field has an effective "inductance" of 0.2 henry at 20 cycles and 0.032 henry at 200 cycles. The total loop function, Figure 7, plots as would an "eddy-current-less" generator with a field inductance of 0.7 henry at all frequencies.

EFFECT OF CARBON-PILE REGULATOR

It is instructive to calculate a simple loop-transfer function with respect to the generated voltage. From this it can be seen easily how the regulator will affect the plot in Figure 7. This calculation for a system with no feedback element Z will be made first.

Let Z_f be the impedance of the shunt-field circuit and let R_o be the steady-state value of carbon-pile resistance. Figure 8 is a plot of measured values of $1/(Z_f+R_o)$. Also let

I_o = the direct component of current flowing through field and carbon pile

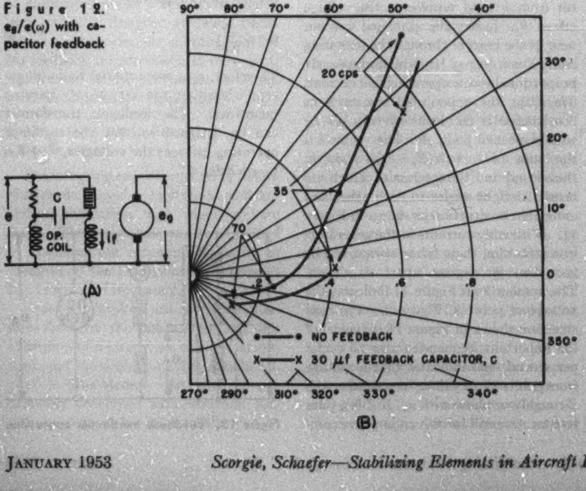
r = the change of carbon-pile resistance from its steady-state value

$H_r=r/e(\omega)$ where H_r is known as the regulator-transfer function

Furthermore, let the transfer function e_g/i_f plotted in Figure 5, be called H_e, so that $e_g=H_ei_f$. The loop-transfer function is derived as follows

$$e=I_or+(Z_f+R_o)i_f=I_oH_re+(Z_f+R_o)i_f \quad (1)$$

Figure 12. $e_g/e(\omega)$ with capacitor feedback



(A)

(B)

$$\frac{i_f}{e}=\frac{1-I_oH_r}{Z_f+R_o} \quad (2)$$

$$\frac{e_g}{e}=(1-I_oH_r)\frac{H_e}{Z_f+R_o} \quad (3)$$

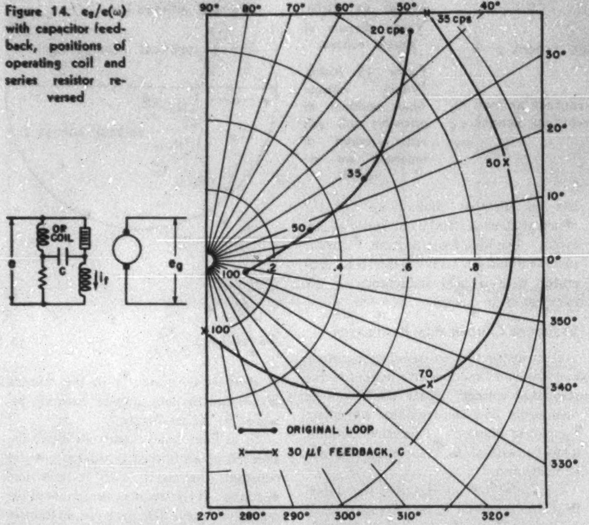
If there is no regulator in the system, $H_r=0$ and the loop-transfer function becomes H_e/Z_f, see Figure 7.

To achieve good steady-state regulation, H_r at $\omega=0$ must be a large positive constant, that is, $1-I_oH_r$ is large and negative. If the frequency response of the regulator were flat over an extremely wide band, the pile-resistance variations would be in phase with the applied voltage for all these frequencies. The quantity $1-I_oH_r$ would remain a negative constant and e_g/e would appear as in Figure 9. No matter what the magnitude of H_r, it would be impossible to have any instability. Phase lags greater than those of Figure 9 can only be attributed to the regulator. In an aircraft regulator there are two causes for r lagging e. One is the effective inductance of the operating coil, while the other cause is the mass of the armature. As the effect of the latter does not become appreciable until around 300 cycles per second, the regulator-transfer function (when no stabilizing elements are added) can be considered to have the form of a simple time lag for frequencies of importance.

Figure 6 illustrates two measured loop-transfer functions with a standard regulator operating and no additional feedback elements. The regulator was adjusted to give better steady-state regulation in curve *b;* that is, the loop has a greater magnitude of response at $\omega=0$. In order to achieve this better regulation, it was necessary to sacrifice speed of transient recovery as is seen from the fact that curve *b* passes closer to the $1+j0$ point. If coils are added to the regulator or resonant circuits put in series with it, the regulator transfer function will be modified as shown in the following section.

If a Z feedback is added to the system, the foregoing analysis breaks down. Under many circumstances, however, it

Figure 14. $e_g/e(\omega)$ with capacitor feedback, positions of operating coil and series resistor reversed



can be considered that the Z will affect only H_r. This assumption is made in the following section. Appendix II shows that for frequencies of interest equation 3 can be approximated by

$$\frac{e_g}{e}=I_oH_r\left(\frac{H_e}{Z_o+R_e}\right) \tag{4}$$

for systems without a Z feedback and for systems with a Z feedback where the preceding assumption is made.

Effect of Resistor Feedback

Figure 10 shows a comparison of the measured loop responses when the impedance Z of Figure 2 becomes a 300-ohm resistance. At $\omega=0$, the loop magnitude has decreased, and correspondingly tests show that the d-c regulation has been damaged. However, the stability has improved. If the regulator-pile screw is then readjusted to regain the good steady-state response, the net effect leaves the loop untouched at $\omega=0$ and yet it passes further from the $1+j0$ point at higher frequencies. At worst, then, the added stability has been gained with only a slight additional power dissipation.

If one assumes no eddy currents in the regulator magnet iron, one can construct a vector diagram, Figure 11, of the voltages in the circuit of Figure 2.

It can be seen from Figure 11 that with operating coil and series resistor in their normal positions, the insertion of a feedback resistor has the effect of adding a component of regulator operating coil current 180 degrees out of phase with the applied voltage in the important range of frequencies. Since the total operating coil current ideally lags at 70–80 degrees at these frequencies; this results in a decrease of the phase lag.

To construct Figure 11, the applied voltage e is laid down as the basic reference. If the impedance Z is large, then the dotted semicircle, with e as a diameter, represents the locus of voltages between point 1 and B, Figure 2. The vector from 1 to A represents the voltage across R_e; hence, the principal component of the current through the operating coil. The voltage $I_or(t)$ is instantaneously proportional to the operating coil current. Therefore, the upper solid semicircle as constructed is the locus of voltages I_or with respect to point A. The vector e is the sum of $I_or+i_fx(R_o+R_f)+N_fd\phi/dt$, the second and third terms of which are nearly at right angles to each other and enter the construction as shown in Figure 11. With eddy currents in the generator-iron structure, these terms always will be less than 90 degrees apart as shown. The number 2 on Figure 11 indicates the voltage of point 2, Figure 2. The construction shown in Figure 11 is typical of the important frequencies near 25 cycles per second, and indicates that a voltage from points 1 to 2 will be impressed across Z roughly in phase with e. If Z is a pure resistor, this will furnish an in-phase component of current through the feedback path. If Z is a capacitor, the proper phasing is disturbed as shown in the measurements, Figure 12.

When the relative positions of the operating coil and series resistor R_e, are reversed, the vector I_or must be referred to point B resulting in the construction of the lower solid semicircle of Figure 13. The voltage 2′ to 1 is impressed across Z under these conditions. This voltage is about 90 degrees out of phase with e, giving the wrong effect.

Initially, it had been hoped to gain some advantage through this maneuver, for the voltages across the feedback are larger with the operating coil circuit inverted in this fashion. This, together with the fact that a capacitor would provide a low impedance and swing the feedback current more nearly in phase with e at the higher frequencies, led to exploration of this means of obtaining greater feedback effects. The failure of the attempt is evident from Figure 14.

Effect of Transformer Feedback

A more effective means of feedback has been mentioned in the literature,[2] namely a transformer connected as shown in Figure 15. To a first approximation, this circuit adds to the regulator operating coil a signal proportional to de/dt when no disturbing input is present in the generator armature. Within the generator, the variation of shunt field voltage e_f, by empirical evidence similar to Figure 7, is known to be related to e_g by the equation

$$\frac{e_g}{e_f}=K_1\left(\frac{\alpha}{j\omega+\alpha}\right) \tag{5}$$

For frequencies where $\omega>\alpha$, the equation becomes approximately

$$e_f=e_g\frac{j\omega}{K_1\alpha} \tag{6}$$

Therefore, e_f is proportional to de_g/dt except when they are very slowly varying quantities. The feedback transformer can be designed so that the regulator operating coil sees the voltage $e_1=e+Ke_f$ $(=e+Kde_g/dt)$.

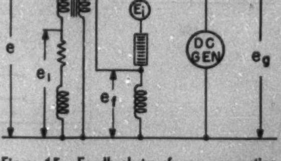
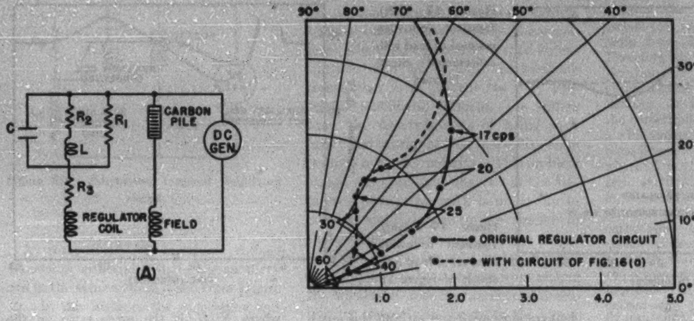


Figure 15. Feedback transformer connection



(A)

(B)

Figure 16. $e_g/e(\omega)$ with antiresonant circuit in operating coil circuit

The characteristic modes of oscillation of the closed system are independent of which quantities are the input and output. Consider input disturbance E_t introduced in the generator field circuit at the position of the carbon pile. Under these conditions in the closed-loop system $e=e_g$ so that $e_1=e+K_1(de)/(dt)$. Also $e_1=rI_o+E_t$. The open loop-transfer function is the product of two terms

$$\frac{e_1}{e_r}=K_g\left(\frac{K_1s+1}{T_gs+1}\right) \tag{7}$$

$$\frac{rI_o}{e_1}=K_r\left(\frac{1}{T_rs+1}\right) \tag{8}$$

If $K_1=T_r$, the loop-transfer function reduces to

$$\frac{rI_o}{e_1}\frac{e_1}{e_r}=K_rK_g\left(\frac{1}{T_gs+1}\right) \tag{9}$$

In other words the phase lags due to the effective inductance of the regulator can be eliminated above a certain frequency by a properly designed feedback transformer. The principle disadvantage of the feedback transformer is in its weight, which must be large in order to obtain good response at low frequencies.

EXPERIMENTS WITH RESONANT STABILIZERS

The foregoing experiments serve to emphasize the physical limitations and analytical difficulties with which we are faced. The design task can be simplified by confining our attention primarily to changing of the loop gain characteristics as a function of frequency.

Bode[4] has worked out many relationships between the phase shift and amplification of a large class of physical systems, and has shown how they may be applied to filter design. Although he was primarily concerned with electronic amplifiers, some of his relationships can be used to show a clear picture of the way in which the phase shift and amplification of our regulator generator control loop behave.

The nature of our system demands that the loop gain A be large at $\omega=0$ in order to achieve good regulation. It is also known that the loop gain falls off rapidly with frequency. If the gain decreases too rapidly, the loop phase will exceed 180 degrees in the critical frequency region and thus produce instability. One relationship between phase and rate of change of A, in decibels with frequency is

$$B(\omega_1)=\frac{\pi}{12}\left(\frac{dA}{du}\right)_{\omega_1}+\frac{1}{6\pi}\int_{-\infty}^{\infty}\left[\frac{dA}{du}-\left(\frac{dA}{du}\right)_{\omega_1}\right]ln\coth\frac{|u|}{2}\,du \tag{10}$$

where

$$u=ln\left(\frac{\omega}{\omega_1}\right)$$

The primary factor determining the phase $B(\omega_1)$ is the slope $(dA)/(du)$ evaluated at ω_1. If the slope everywhere were the same, the integral in equation 7 would vanish. This is never true in physical systems, and the departure of $(dA)/(du)$ at all frequencies from its value at ω_1 will influence equation 10. The weighting factor, $ln\coth(|u|)/2$, rises to a very large peak near the frequency $\omega=\omega_1$, hence, lends very large weight to the integral of equation 10 primarily in that region. At frequencies fairly remote from ω_1 the weighting factor becomes approximately $(2\omega)/\omega_1$, for $\omega<<\omega_1$, and approximately $(2\omega_1)/\omega$, for $\omega_1<<\omega$. For the purpose of the discussion which follows ω_c will be defined as the frequency at which each loop comes closest to the $1+j0$ point. We will be concerned with obtaining the best possible phase margin at ω_c by decreasing $(dA)/(du)$ in the frequency band near ω_c. The frequency ω_c does not occur at any fixed loop gain, but except for extraordinary cases falls in the region where the loop gain is between 0 and 1.0.

The regulator-time constant can be decreased greatly by using fewer magnet turns of greater current carrying capacity, in which case ω_c is raised and the corresponding loop gain is lowered. This is due to the fact that the average magnitude of $(dA)/(du)$ at lower frequencies has been decreased, thus increasing the phase margin.

If however, we do not wish to dissipate the extra d-c power, the loop gain curve can be shaped differently. Starting with the original operating coil, a cascade circuit can be added as shown in Figure 16 (A), which will not affect the d-c regulator current. That is to say, $A(=20ln|e_g/e|_{\omega=0})$ remains unchanged. At a sufficiently high frequency the current in both configurations will be determined primarily by the effective inductance of the regulator operating coil. The frequency at which this is true will be called ω_∞ and the corresponding loop gain and phase A_∞, B_∞. The loop phase B over the entire frequency range is related to these gains by

$$\int_0^{\omega_\infty}\frac{B}{\omega}\,d\omega=\frac{\pi}{2}(A_\infty-A_0) \tag{11}$$

A similar useful relationship is

$$-\frac{\pi}{2}B_\infty=\int^{\omega_\infty}(A-A_\infty)d\omega \tag{12}$$

which can be written

$$\int_0^{\omega_c}A\,d\omega+\int_{\omega_c}^{\omega_\infty}A\,d\omega=\text{a constant} \tag{13}$$

The objective is to introduce a rapid decrease in loop gain at low frequencies where the weighting factor of equation 10 is $(2\omega)/\omega_c$, which is relatively insignificant. The first integral of equation 13 will then have been decreased. It follows

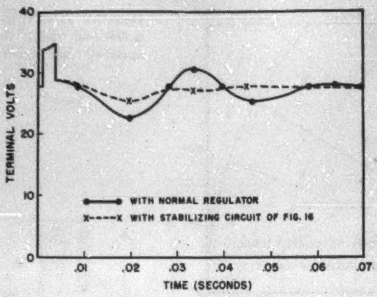


Figure 17 (left). Transient following removal of load with antiresonant circuit inserted

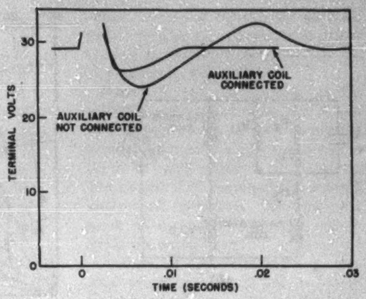


Figure 19 (right). Transient following removal of load with circuit of Figure 18

that the second integral of equation 13 must necessarily increase and as a result it is highly probable that $(dA)/(du)$ in the region near ω_c will have been reduced. From equation 10 it follows that the phase margin at ω_c will have been improved. This, indeed, is the case as is seen from Figure 16(B). The original ω_c corresponded to a frequency of 35 cycles per second, and L and C were chosen so as to resonate at 19 cycles per second. The average slope, $(dA)/(du)$, through the frequency range 20–40 cycles per second has been roughly halved. From equations 10 and 11 it can be seen that in order to decrease the loop phase lag in the critical frequency range, the phase lag somewhere outside this range must necessarily have increased. However, this increase occurs primarily at very low frequencies where the net effect is insignificant. Best design normally demands the use of an electrolytic capacitance, which formerly would have militated against the idea. However, recently developed tantalum electrolytic capacitors meet the low temperature requirements in aircraft and are characterized by low leakage currents and high reliability. A comparison of transient response with and without this auxiliary circuit is shown in Figure 17.

The principal difficulty with this design lies in adjusting the temperature changes in the copper winding of the inductor. A design which eliminates the temperature change difficulty, and, incidentally, the inductor, but which requires a redesigned regulator, is shown in Figure 18(A). The inductance is wound directly on the magnet circuit of the regulator and connected in series with a capacitor C across the bus terminals. The auxiliary coil must be wound so that current flowing from the positive to the negative terminals will aid the main magnetomotive force. No direct current flows in the auxiliary coil; consequently, the temperature of the winding does not affect regulation.

The transfer function of the regulator shown in Figure 18(A) is found to be

$$\frac{r}{e}=K_r\times\left[\frac{sC\left(R_1+R_2\dfrac{N_1}{N_2}\right)+1}{s^2C\dfrac{L_2}{R_2}\left(R_1+R_2\dfrac{L_1}{L_2}\right)+sC\left(R_1+\dfrac{L_2}{CR_2}\right)+1}\right] \tag{14}$$

where N_1/N_2 is the turns ratio on the magnet core. The auxiliary winding has added one degree of s to numerator and denominator, leaving B_∞ unchanged. If R_1 is very large, equation 14 reduces to the original simple time lag with time constant L_2/R_2. If, however, R_1 is small, the system performance must improve if

$$\frac{L_2}{R_2}\geqq C\left(R_1+R_2\frac{N_1}{N_2}\right)>C\sqrt{\left(R_1+R_2\frac{L_1}{L_2}\right)\frac{L_2}{R_2}} \tag{15}$$

If R_1 is negligibly small, this reduces to

$$\frac{L_2}{R_2}\geqq CR_2\frac{N_1}{N_2}>\frac{N_1}{N_2}\sqrt{CL_2} \tag{16}$$

R_2 and L_2 are fixed by power dissipation consideration, spring and magnet characteristics, and so forth. The second inequality in equation 16 can, therefore, be satisfied by making C sufficiently large. Once this is done, the 1st inequality of equation 16 can be satisfied by assuring N_1/N_2 does not become too great.

Figure 18(B) shows a comparison of measured transfer functions with and without the auxiliary coil of 18(A) connected.

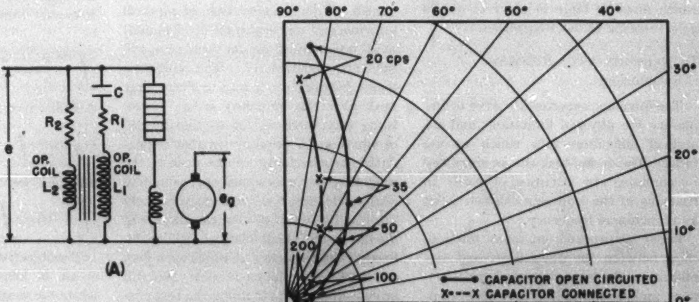


Figure 18. $e_g/e(\omega)$ with auxiliary operating coil in resonant circuit

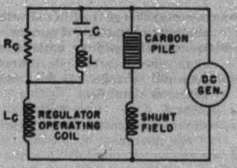


Figure 20. Alternative resonant stabilizing circuit

A comparison of terminal voltage transients with and without the auxiliary coil in the same system is shown in Figure 19. In this instance the experiment was performed using a standard regulator whose main operating coil was partially removed to make room for the auxiliary coil winding. The remainder of the main coil, of course, had to be overloaded for the test. Its inductance was decreased, which in itself had a stabilizing effect upon the system, so that no direct comparison of test results with previous experiments was possible.

The second design, with an auxiliary *CL*, operates on a slightly different assumption, although the basic reasoning is the same. Each is designed to decrease the magnitude of $(dA)/(du)$ in the vicinity of ω_c. However, in the second case there is no frequency below ω_c where the $(dA)/(du)$ has been increased beyond its initial value. The objective here is to decrease $(dA)/(du)$ at lower frequencies and postpone the large rate of change for some higher frequency. Equations 11, 12, and 13 cannot be used to analyze this circuit because of the fact that A_∞ changes with the addition of the auxiliary magnet winding.

Parenthetically, it should be stated that the circuit of Figure 20 operates in a fashion roughly the equivalent of that in Figure 18(A). The main question is where to wind the auxiliary inductance. Here equations 10 through 13 all apply.

Summary

The function and design of feedback and cascade components in regulated aircraft generator systems has been discussed in detail. A method for measuring frequency response characteristics of the control loop with both carbon-pile voltage regulator and feedback elements has been shown.

The principal importance of the work lies in the background which it provides for extension to more complicated systems, for at the present time stability problems in d-c systems have been overcome fairly well. The worst similar problem facing designers of aircraft electric systems today is the stability of a-c systems, where a d-c generator is used for excitation. For a-c systems the feedback transformer is at present the only successful stabilizing element. It suffers from the disadvantages of being very heavy and causing serious instability upon reversal of exciter polarity. This problem also is susceptible of analysis in the frequency plane, and work is underway to provide stabilizing networks similar to those discussed here.

In this study of the d-c system a series of examples has been chosen from among the successful and the unsuccessful stabilizing elements to illustrate the agreement between theory and practice and to acquire a feeling for the limitations on what might be accomplished with a given generator. It has been shown that the most advantageous impedance *Z* Figure 2, is a simple resistor.

Large stabilizing effects in relatively narrow frequency bands are possible with resonant circuits. Arrangements for accomplishing these effects are numerous and several of the simplest have been illustrated here. Their design is always restricted by the basic physical relationships between phase and magnitude of the loop frequency response.

Measurements have shown that eddy currents in typical aircraft generators have a very appreciable detrimental effect upon system stability. Further experimental studies of eddy-current effect are being carried out.

Appendix I. Experimental Technique

Our objective in the experimental work was to measure the phase and magnitude relations of the small sinusoidal variations, e and i_f, also i_f and e_g, for a band of frequencies approximately from 0 to 200 cycles per second. A concomitant objective was to supply the normal, stable background direct current upon which to superimpose these variations. The regulator characteristic is such that if one simply keeps increasing the operating coil voltage, the carbon-pile resistance will discontinuously jump from some low value to infinity.[2] Therefore, to measure i_f/e, a feedback device must be used which will directly and automatically control the carbon-pile resistance. In the configuration of Figure 2, the carbon-pile current I can vary through a wide range, hence the feedback controller must be made sensitive to the quotient E/I, where E is the voltage across the carbon pile. A circuit to accomplish this is shown in Figure 21. The principal element is a photoformer, which is comprised of a d-c cathode-ray oscilloscope with part of its screen blacked out, and a photomultiplier to detect variations of light intensity from the screen of the scope. The voltage E is impressed on the vertical plates of the cathode-ray tube and a voltage proportional to I is impressed across the horizontal plates. Diagonally across the face of the cathode-ray tube a straight line is drawn, below which everything is blacked out. If the carbon-pile resistance (the ratio of E/I) is low the dot cannot be seen by the photomultiplier and the *6AS7* grids have no bias. If the resistance is high the dot can be seen by the photomultiplier, and a 70-volt negative bias is applied to the *6AS7* grids. When the resistance is at its proper value, the dot is sitting on the line with enough of its area showing to put the proper bias on the *6AS7*'s. The steady-state value of the carbon-pile resistance can be varied by adjusting the mask position and/or adjusting the sensitivity of the x or y axis amplifier. The oscillator shown in Figure 21 superimposes the required sinusoidal variations upon the steady-state background. Stability considerations demand that the voltage sensing plates of the scope be connected across both the field and the carbon pile in series.

When the circuit is in operation, two undesirable signals are introduced in the system. The first is thermal noise from the photomultiplier, the majority of which is at an exceedingly high frequency, although some components exist below 200 cycles. The second spurious voltage appears as harmonics of the oscillator signal. These harmonics are introduced in the photoformer rather than in the circuitry under measurement. From other experiments it is known that the system under measurement is linear for the variations imposed upon it. Therefore e and i_f were filtered before measurement using the same band-pass filter for each. Their magnitudes were measured with a vacuum-tube voltmeter and their phases were both measured with respect to that of the input oscillator.

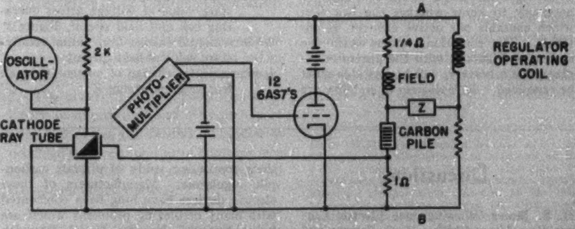


Figure 21. Electronic circuit for measuring transfer functions

For the measurement of e_g/i_f a sinusoidal current (superimposed on the proper direct current) of the same order of magnitude as flowed during the measurements of i_f/e was caused to flow through the field. The generator was run and the output voltage variations noted. As commutator ripple caused a great amount of noise, filtering was employed. It was found that this noise got considerably worse, with less impedance than that of a scope connected across the terminals, that is, a moving coil voltmeter across the terminals caused the noise to increase drastically. It was found that these measurements could be checked, with noise eliminated, by use of a search coil on the pole face. This function e_g/i_f shown in Figure 5 varied somewhat with the magnitude of i_f, but these variations were not great. As has been stated, the loop-transfer functions were obtained by multiplying e_g/i_f by the various measured values of the ratio i_f/e.

Appendix II. Choice of Open Loops

The closed system being considered (see Figure 1) contains a feedback impedance. Consequently, the regulator characteristics cannot be examined as a unique box separate from the generator. Yet it is desirable, because of simplifications in measurement and analysis, to split the system into independent boxes which can be examined alone and fitted into convenient block diagram similar to those used in servomechanism analysis.[5] It is also desirable to be able to open the system at a suitable point, insert a disturbance signal, and evaluate the response of a complete open loop. In a system such as the one we are considering, where there are several active elements and appreciable loading among the various mesh currents, the open-loop transfer function will be different depending upon where the system was broken.

When the open-loop transfer function with respect to the generator is desired, the transfer impedance e_g/i_f of the generator is taken as zero and an independent generator e inserted in its place. The relation i_f/e under these circumstances constitutes one box of the loop and can be multiplied by e_g/i_f to give the total loop transfer function. System instability will result if $1-(e_g/i_f)(i_f/e)$ has a zero in the right half plane. The simple Nyquist criterion applied to the open loop is not a definitive test unless one knows that the loop has no poles in the right half plane. That is to say it is necessary to know whether the box i_f/e, which contains one active element, is unstable. The conditions under which the system of Figure 1, with the generator replaced by a battery, will be unstable must be examined. Since degrees of stability are of no interest, the application of Routh's criterion to the characteristic determinant is sufficient.

The open-loop transfer function with respect to the generator was chosen in this paper only due to experimental convenience. In a previous paper, it was expedient to choose the open-loop transfer function with respect to the regulator. To utilize the Nyquist stability criterion with respect to the loop, it is sufficient to know that the slope of the generator saturation curve, at the vicinity of the operating point, is smaller than the resistance of the shunt field circuit.

In practical systems the loop-transfer function with respect to both the regulator and generator are very nearly the same in the vicinity of the $1+j0$ point. When the job is to improve an already stable system, it is useful to simplify these loops in the critical frequency region.

To a first approximation the variation of voltage across the carbon-pile resistor is given by

$$I_0 r = I_0 e\left(K_r \frac{\alpha_r}{s+\alpha_r}\right) \tag{17}$$

where K_r typically is about 15, α_r is on the order of 70, and I_0 is on the order of 1, or greater. The critical frequencies occur near $\omega = 2\alpha_r$ so that $I_0 r/e > 7$. At the critical frequency $e > e_g$. Therefore $I_0 r$ is far greater than both e_g and e in the critical region, so that the generator field circuit voltage variation is approximately equal to $I_0 r$. The field current variation must then equal approximately

$$i_f = \frac{I_0 r}{R_0 + Z_f} \tag{18}$$

The loop-transfer function $r/e \times i_f/r \times e_g/i_f$ then is given by

$$\frac{e_g}{e} = H_r\left(\frac{I_0}{R_0+Z_f}\right)H_g \tag{19}$$

(that is, equation 3 with 1 considered negligible) in the neighborhood of the critical frequency with or without the feedback impedance Z. The presence of H_r as a multiplying factor in equation 19 makes it possible to concern ourselves only with variations of this quantity when examining Z feedback.

Appendix III. Nomenclature

e_g = incremental change of generated voltage from initial steady-state value
e = incremental change from initial steady-state value of voltage across operating coil and field circuit, Figure 2
i_f = incremental change from initial steady-state value of field current
i_b = incremental change of current through feedback impedance Z
i_1 = incremental change from initial value of regulator operating coil current
r = incremental change from initial steady-state value of carbon-pile resistance
ϕ_f = incremental change of flux linking generator shunt field
$e_r = I_0 r$, where
  I_0 = direct component of total carbon-pile current
  $k = e_r/i_1$
  $H_r = r/e$
  $H_g = e_g/i_f$
Z = impedance of passive feedback elements connected as shown in Figure 1
Z_f = impedance of generator shunt field
R_b = the value of Z when it is a resistor
R_c = resistance of regulator operating coil circuit
R_0 = steady-state value of carbon-pile resistance
L_c = equivalent inductance of regulator operating coil circuit
R_1, R_2, L_1, L_2 = resistors and inductances indicated in Figure 18(A).

$$K = \frac{R_0 + k}{R_b + R_c}$$

$$T = \frac{L_c}{R_c}$$

s = the Laplace operator
N_f = number of turns on generator shunt field

$$u = ln\left(\frac{\omega}{\omega_1}\right)$$

ω_c = the frequency at which a particular loop-transfer function approaches closest to the $1+j0$ point
ω_∞ = a frequency sufficiently high, if one exists, above which the loop behavior is independent of the auxiliary circuit whose effect is being evaluated
$B(\omega)$ = the loop phase as a function of frequency
$A(\omega)$ = loop gain in decibels $\left(=20\, ln\left|\frac{e_g}{e}\right|\right)$
$A_0 = A$ at $\omega = 0$
A_∞, B_∞ = A and B evaluated at ω_∞

References

1. Dynamic Characteristics of Carbon Pile Voltage Regulators, **R. L. Mills**. *AIEE Transactions*, volume 69, part II, 1950, pages 1347–53.

2. Transient Analysis of a Voltage-Regulated Aircraft D-C System, **Donald G. Scorgie**. *AIEE Transactions*, volume 69, part II, 1950, pages 1318–23.

3. Instability in D-C Aircraft Systems, **H. B. Bunce, J. C. Cunningham, W. M. Davidson**. *Electrical Engineering*, volume 64, November 1945, pages 786–90.

4. Network Analysis and Feedback Amplifier Design (book), **H. W. Bode**. D. Van Nostrand Company (New York, N.Y.), 1945.

5. Theory of Servomechanisms (book), **Hubert M. James, Nathaniel B. Nichols, Ralph S. Phillips**. McGraw-Hill Book Company, Inc. (New York, N. Y.), 1947.

Discussion

H. B. James (Westinghouse Electric Corporation, Lima, Ohio): D. G. Scorgie and D. H. Schaefer should be complimented on their continuing study of aircraft carbon-pile regulators. Manufacturers of these electric devices have long been confronted with many perplexing problems which are being explored by the Naval Research Laboratory. We can all benefit from the results of these investigations. Such facts as were discussed in their last paper[1] pertaining to the effect of hysteresis of the magnetic structure and their present findings of significant eddy-current effects in the generator have added valuable information to both the study of regulator stability requirements and a more thorough understanding of the associated rotating equipment.

Especially gratifying is the application of "standard" servomechanism theory to this particular field. Such treatment has been noticeably absent in past literature.

The authors do make one statement in this paper which should be commented on. They say that "for a-c systems, the feedback transformer is at present the only successful stabilizing combination." Successful stabilization has been achieved by resistance-capacitance networks feeding a separate coil on the regulator magnet and by the use of the generator field voltage supplying a signal to a separate coil which has an additive polarity with respect to the main regulator coil (a third coil, responsive to generator field current, is connected with opposing polarity to eliminate effects on the regulation).

REFERENCE

1. STEADY-STATE CHARACTERISTICS OF CARBON-PILE VOLTAGE REGULATORS, D. G. Scorgie, D. H. Schaefer. *AIEE Transactions*, volume 70, part II, 1951, pages 1572–77.

D. C. Scorgie and **D. H. Schaefer:** The authors thank Mr. James both for his kind words and for the information contained in his discussion. It might be noted that there is now still another available way of stabilizing a-c systems. This is by energizing the regulator operating coil from the lead network described in another paper by Mr. Scorgie.[1]

REFERENCE

1. LEAD NETWORKS UTILIZING A SATURABLE CORE MEMORY, D. G. Scorgie. *AIEE Transactions* volume 71, part II, November 1952, pages 297–302

Electrostatic Oscillation in Sensitive Contact Mechanisms

THOMAS R. STUELPNAGEL
ASSOCIATE MEMBER AIEE

Synopsis: Destructive contact "frying" is particularly troublesome in sensitive contact mechanisms such as thermostats, small switches, and polarized relays. This paper describes an investigation which proved that sensitive contact frying is actually an electromechanical oscillation of the contacts and is caused principally by electrostatic attraction between the contacts. On the basis of this analysis, simple changes in contact shapes and detent arrangements are proposed as a means of efficiently eliminating the phenomenon and thus improving sensitive contact performance.

DEMANDS are being made continuously for improved operation in sensitive contact devices such as thermostats, switches, and polarized relays. One of the important problems in the design of such contact mechanisms is the elimination of destructive contact oscillation or frying, as it is commonly called. This disturbance is particularly troublesome in sensitive contact mechanisms where contact movement is proportional to the operating signal. Oscillation occurs just before opening or after closing when the physical separation between the contacts is small.

In ordinary switch mechanisms, such as solenoid relays and toggle switches, the problem of what happens when contacts are brought in close proximity to each other or closed with almost zero contact pressure is sidestepped by making provisions for a rapid make and break. Magnetic or mechanical detent in one form or another is employed. However, when dealing with a sensitive switch device this problem must be met squarely because the loss in on-off sensitivity resulting from a generous amount of detent often cannot be tolerated.

For example: In thermostats detent is undesirable because it sets the tolerance on the temperature being controlled. In a sensitive relay used as a servomechanism sensing element, excessive detent can result in system oscillation. In snap-action switches the mechanical detent determines the amount of plunger travel required between make and break.

The most efficient way to prevent frying in sensitive contacts is to minimize the contact oscillation forces first and then supply only enough detent to overcome the remaining oscillation forces. Determination of how to minimize these oscillation forces, however, requires an investigation of the frying phenomenon. This paper discusses such an investigation. Improvement of the sensitive polarized relay shown in Figure 1 was the object of this study; however, the results obtained are applicable to sensitive contacts in general.

Characteristics of Contact Oscillation

Contact frying was found to be a serious problem in the sensitive relay as used in the 120-volt d-c control circuit of Figure 2. The observed phenomenon was oscillatory in nature and had a frequency of over 200 cycles per second. Oscillation occurred when either pair of contacts was closed more than 90 per cent of the full 0.006-inch contact travel.

An attempt was first made to stabilize the contacts of this relay by simply adding more detent. It was found, however, that the amount of detent required resulted in a considerable loss in relay on-off sensitivity and, consequently, servo sensitivity. As an alternate approach the oscillation phenomenon was investigated.

Experimental Isolation of Contact Frying

Analysis of the observed oscillation was complicated by the small forces and distances involved. It was necessary to measure contact forces of less than 1 gram and movements of less than 1,000th inch. In the process of isolating all possible sources of oscillation the signal coil and permanent magnetic field of the test relay were removed so that only the contact assembly remained. The relay was then shock-mounted in a chamber which could be evacuated. This arrangement excluded the possibility of magnetic or mechanical feedback and any effect from the atmosphere.

The contact assembly was operated in a vacuum to prove that minute thermal expansions in the air between the contacts were not contributing to the contact oscillation. It was thought that such a disturbance could result from localized heating of the contact surfaces at the instant of electric contact.

The contacts were mechanically positioned in the unstable region (within 1/2,000th inch of each other) so that oscillation would occur. Then the circuit constants and armature mass were varied while the contact current and voltage were monitored simultaneously on a dual-beam oscilloscope. After establishing the nature of the oscillation the test relay was

---

Paper 52-328, recommended by the AIEE Air Transportation Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Middle Eastern District Meeting, Toledo, Ohio, October 28–30, 1952. Manuscript submitted October 25, 1951; made available for printing September 5, 1952.

THOMAS R. STUELPNAGEL is at 3922 San Anseline Avenue, Long Beach, Calif.

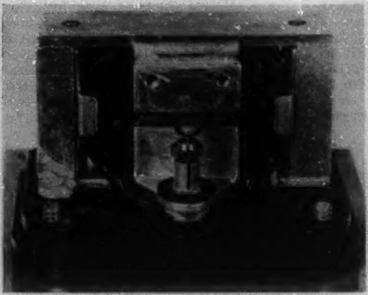
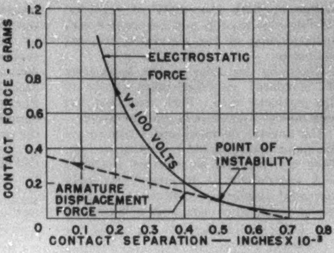


Figure 1 (left). A polarized sensitive relay with contacts reworked to reduce electrostatic oscillation (Barber-Colman Micropositioner)

Figure 3 (right). Calculated electrostatic attraction between flat electric contacts of 0.096-inch diameter and operating at 100 volts d-c



reassembled and the actual oscillation region measured.

Electrostatic attraction between the contacts of the relay was found to be the controlling cause of the oscillation.

Mechanism of Electrostatic Oscillation

The electrostatic attraction force between two oppositely charged flat contacts in air is shown in equation 1 to be proportional to the surface area and the square of the voltage, and inversely proportional to the square of the contact separation distance. The frequency of contact oscillation resulting from the electrostatic attraction is dependent upon the mass and spring constant of the contact armature and the rate of voltage build-up across the contacts after separation.

$$F=4.5\times10^{-10}(V^2A/X^2) \qquad (1)$$[5]

where

F = attraction force, grams
V = voltage across contacts, volts
A = effective contact area, inches squared
X = contact separation, inches

THEORETICAL EXAMPLE OF ELECTROSTATIC OSCILLATION

The theoretical curve presented in Figure 3 perhaps best describes the mechanics of electrostatic contact oscillation. A contact voltage of 100 volts d-c and flat contacts with a diameter of 0.096 inch has been assumed in this calculation. Superimposed on the electrostatic force curve is the displacement curve of a linear armature having a spring constant equal to 1/2 gram per 1,000th inch. From this diagram it can be seen that the relay contacts become unstable when the electrostatic force curve balances the armature displacement force curve. This condition occurs in this example when the contacts are positioned to within about 1/2,000th inch of each other. Any additional signal to the sensitive coil or slight vibration of the relay causes the electrostatic force to override the armature spring and the contacts are pulled together. Once the contacts electrically close, the voltage and consequently the electrostatic force between the contacts drops to zero. If detent in the form of a microscopic weld or magnetic coupling from the load current is not sufficient to hold the armature closed against the armature displacement force of 0.35 gram, the contacts separate, the voltage builds up, and the cycle repeats at some frequency and amplitude determined by the physical armature constants and the rate of voltage build-up.

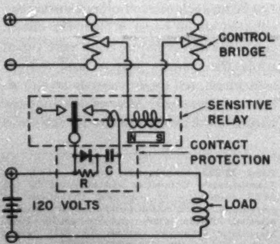


Figure 2. A polarized sensitive relay as used in a resistance bridge control circuit

ACTUAL EXAMPLE OF ELECTROSTATIC OSCILLATION

An oscillogram of electrostatic oscillation with physical and electrical conditions of the test relay closely representing the calculated curve is shown in Figure 4. The complete circuit, Figure 2, consists of a 120-volt d-c supply voltage, a polarized relay adjusted to a 6-ampere-turn sensitivity at a 0.006-inch contact gap, a resistance-capacity-diode contact protection network, and a relay coil load.

From this oscillogram it can be seen that the contact protection capacitor discharges a peak current of 75 milliamperes through the contacts at the instant of closure. After remaining closed for about 1 millisecond the contacts separate. The transient current of 25 milliamperes, flowing in the inductive load, discharges into the contact protection capacitor, thereby raising the voltage momentarily to 200 volts d-c before the contacts are drawn closed on the second oscillatory cycle. The frequency of oscillation in this case is seen to be 230 cycles per second. Unless a small contact weld forms, this oscillation sustains until the contacts are repositioned outside the critical region or until microscopic protrusions form on the contact surface which prevent the contacts from closing completely, thus reducing the electrostatic force. It was not unusual for oscillations to sustain for 10 minutes in the circuit of Figure 2.

MEASUREMENT OF CONTACT OSCILLATION

The oscillation region was measured by adjusting the test relay to close at a given signal current with a fraction of a volt across the contacts. Increments of voltage up to 200 volts d-c then were connected to the contacts and the difference measured in the amount of signal current required to close the contacts. The signal current was converted to ampere-turns and then to distance by calibrating the movement of the test relay armature with a micrometer microscope. The measured armature hysteresis curve is shown in Figure 5. The critical oscillation region has been superimposed on this curve.

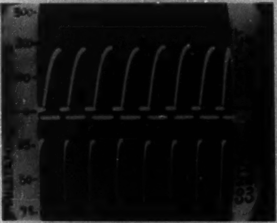


Figure 4. Oscillogram of electrostatic contact oscillation at 120 volts d-c

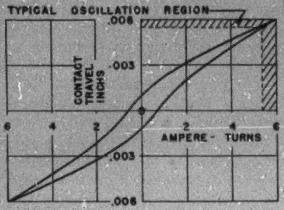


Figure 5 (left). Hysteresis curve of sensitive relay armature showing the critical electrostatic oscillation region

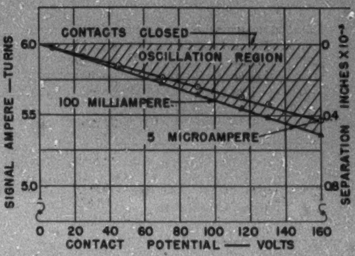


Figure 6 (right). Measured electrostatic attraction between sensitive contacts

Typical data obtained for two values of contact current, 5 microamperes and 100 milliamperes, are presented in Figure 6. These curves prove that the phenomenon is essentially independent of contact current and is caused by electrostatic force. The scale on the right of Figure 6 relates the contact separation to signal ampere-turns.

Electrostatic Oscillation at 24 Volts D-C

The curves of Figure 6 also demonstrate that a linear relationship exists between the points of armature instability and voltage. Although it is probable that the shape of this curve will be dependent upon the configuration of the electrostatic field for different contact surface conditions, it is significant that the oscillation region does not necessarily decrease as the square of the contact voltage. For this reason the electrostatic oscillation problem requires proportionate consideration at lower voltage contact applications such as 24 volts d-c.

As an example of electrostatic oscillation at lower voltages an oscillogram recorded at 24 volts d-c is shown in Figure 7. The test conditions were the same as used in the 120-volt d-c circuit of Figure 2. At the lower voltage the oscillation frequency decreased from 230 to 170 cycles per second but the general characteristic remained the same.

Further Experimental Observations

No attempt was made in this work to co-ordinate closely the measured and calculated points of armature instability because in the practical application large variations in effective contact area can occur each cycle. The magnitude of the effects, however, is shown to agree in Figures 3 and 6 where a point of contact instability at 100 volts d-c was calculated to occur at 0.0005 inch and measured at 0.0004 inch.

Further tests demonstrated the following additional characteristics of electrostatic contact oscillation:

1. Changes in atmospheric pressure from sea level to 29 inches of mercury had no effect on oscillation frequency or amplitude.

2. An increase in armature inertia of several times reduced the oscillation frequency from 230 to 60 cycles per second.

3. Alteration of the time constant of the load or the contact protection circuit had only minor effects on the oscillation frequency of 230 cycles per second.

4. Removal of the contact protection circuit, while switching an inductive load, resulted in contact welding forces sufficient to overcome the electrostatic oscillation.

5. Removal of the contact protection circuit, but with a resistive load substituted for the relay coil, resulted in oscillation frequencies as high as 3,000 cycles per second. The stability of this oscillation decreased as the current was increased to 200 milliamperes because after a few rapid cycles of these higher currents the contacts welded closed with a force sufficient to overcome the armature displacement force. On the other hand, when contact protection was employed, capacitor discharge currents of 1/2 ampere or more were required to provide stabilizing welding forces at the normal oscillation frequency of 230 cycles per second.

Contact Failure Resulting from Electrostatic Oscillation

The observation just given emphasize two reasons why electrostatic oscillation is to be avoided in a sensitive relay installation:

1. If the oscillation sustains indefinitely before and after each relay operation, the contacts are being pulsed continually with useless energy which causes erosion of the contacts themselves and vibration wear in the controlled relay or magnetic clutch. In other applications, such as switching plate power in vacuum-tube circuits, the sustained oscillation can cause complete failure of the installation.

2. The build-up of thermal energy at the microscopic switching surfaces during a series of rapid oscillatory cycles can cause a weld many times stronger than could be caused by a single switch cycle. A contact weld, requiring even a fraction of a gram of rupturing force, can result in erratic performance or possible failure of some sensitive devices.

Elimination of Electrostatic Oscillation

Once the magnitude of the electrostatic oscillation forces in a pair of contacts is recognized, the actual suppression of the oscillation is accomplished most easily first by reducing the physical contact area, and second by applying enough detent to overcome the remaining electrostatic force.

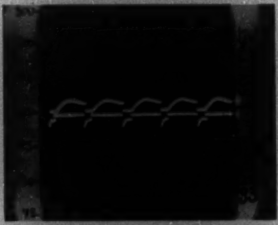


Figure 7 (above). Oscillogram of electrostatic contact oscillation at 24 volts d-c

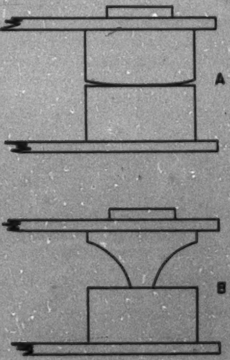


Figure 8 (right). Shaping of typical sensitive contacts to reduce electrostatic attraction

A—original 0.096-inch-diameter contacts
B—reworked 0.020-inch-diameter contact

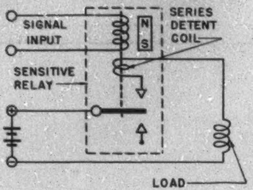


Figure 9. Series detent arrangement for a polarized sensitive relay

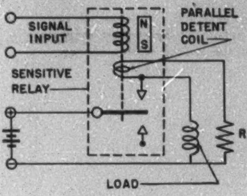


Figure 11. Parallel detent arrangement for a polarized sensitive relay

Reduction of Contact Area

Since for typical contacts the electrostatic force is troublesome at separation distances of less than 1,000th inch, it is only necessary to undercut one of the contacts a few thousandths, leaving a smaller diameter area for actual contact. For example, a 95-per-cent reduction in electrostatic force was obtained in the relay of Figure 1 by reducing one of the original 0.096-inch-diameter flat contacts to 0.020-inch diameter in the region of contact. A sketch of the reworked contact assembly is shown in Figure 8.

The mechanical and electrical stability of the flat 0.020-inch-diameter contact was found to be excellent when a relatively hard material such as platinum alloy was used. In the circuit of Figure 2, no mechanical deformation and negligible metal transfer was measured after 10,000,000 operations of a test contact cycled against a flat stationary contact with a driving force of 30 grams.

Addition of Detent

In addition to reducing the contact area, some detent also is required to overcome the small remaining electrostatic force. Since it is advantageous to select the most practical and efficient type of detent for the particular application, the characteristics of three magnetic detent arrangements, as related to this type of contact stabilization, are discussed in the following. An example also is presented to demonstrate the improved performance which can be obtained in a typical servo control application of the test relay by using a minimum amount of detent.

Methods for Obtaining Contact Detent

Permanent Magnet

One method for adding detent to a proportional contact mechanism is to locate a small permanent magnet at the end of armature travel which will snap the armature closed from some specified distance before actual electric contact has been made. This type of detent has the advantage of being independent of the contact current and consequently works equally well on a-c or d-c circuits. The difficulty of controlling the amount of detent is perhaps the chief disadvantage.

Series Detent Coils

The second method is also magnetic but can be used only with d-c contact currents when incorporated in a polarized relay. Small auxiliary coils, one for each contact of a center-off relay, are wound on the signal coil bobbin, connected to the respective contact circuit, and polarized so as to provide a positive force when the contacts are closed.

A series arrangement of this type of detent, shown in the circuit diagram of Figure 9, is often used in d-c sensitive relays because the detent force can be controlled readily by specifying the number of detent ampere-turns required as related to the ampere-turn sensitivity of the relay. Also this type of detent does not alter the relay sensitivity during contact closure.

The series detent has a serious disadvantage, however, because the detent force is dependent upon the transient conditions of the contact current. Consequently when the contact load is inductive, as it usually is, a constant detent force is not established until several milliseconds after contact closure. Even when a contact protection circuit of the type shown in Figure 2 is used, the detent force will dip between the time that the contact protection circuit discharges and the load current increases. The efficiency of magnetic detent placed in series with an inductive load thus is limited during the first few milliseconds when it is needed to stabilize the contacts. Therefore, in the specification of series detent, consideration must be given to the time constants of the contact protection circuit and the inductive load. In the oscillogram of Figure 10, which shows a nonoscillatory contact operation, the effective detent current is 25 milliamperes during contact closure. Yet at contact opening a sacrifice in differential sensitivity of twice this value results because of the stabilized 50-milliampere current. Contact protection circuits with longer time constants can be used to reduce the seriousness of the current dip; however, in high-speed servo applications, for example, this approach can increase the drop-out time of the controlled relay or clutch an undesirable amount.

Parallel Detent Coils

Increased detent efficiency can be obtained with the parallel detent arrangement described in Figure 11 since a constant detent force is established almost instantaneously after contact closure. This type of detent usually is not justified, however, because of the additional terminal complication, larger number of detent turns, and resistor required. Further-

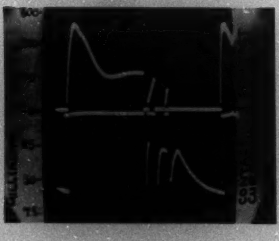


Figure 10 (left). Oscillogram of sensitive contact operation in the circuit of Figure 2

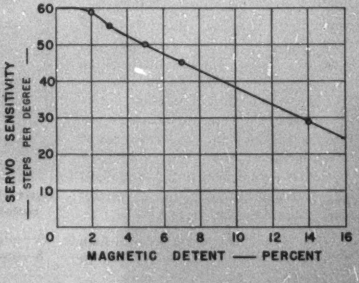


Figure 12 (right). Reduction in servomechanism sensitivity caused by contact detent. Sensitivity is defined as the minimum incremental movement of the servo actuator more, the shunt effect of the parallel circuit, unless considerably higher than the load resistance, will slow down the drop-out time of the controlled slave relay or magnetic clutch. In servo applications this will have the same effect on system stability as excessive detent.

EFFECT OF DETENT ON SERVO PERFORMANCE

An example of the effect of detent on the performance of a resistance bridge controlled servomechanism is presented in Figure 12. In this example it can be seen that a detent force of 15 per cent of a 6-ampere-turn 1-milliwatt sensing relay reduced servo step sensitivity by more than 50 per cent. Since step sensitivity of a nonlinear servo is a criterion of system stability, it is apparent that minimum detent in the sensing relay is a problem requiring investigation.

SELECTION OF OPTIMUM DETENT

In the test relay application it was found that $1^{1}/_{2}$ per cent parallel detent or 3 per cent series detent was needed for contact stability with the reduced contact area. A series detent force of twice the minimum value, or 6 per cent, actually was used. With this arrangement it was possible to combine optimum servomechanism operation with contact stability and without reducing the reliability of the basic relay.

Conclusions

Electrostatic attraction between the contacts of a sensitive switch or relay has been found to be responsible for much of the contact oscillation or frying observed in these devices. Electrostatic force, generally considered as an insignificant factor in contact operation, becomes important when dealing with the gram forces and milli-inch distances which characterize sensitive contact devices.

The effects of electrostatic contact attraction for a given voltage can be minimized by physically shaping the contact surfaces. In this way stable sensitive contact operation can be obtained without the use of excessive detent. Optimum differential sensitivity, so desirable in sensitive contact mechanisms, results.

Even after reducing the effect of electrostatic contact attraction, some detent is needed to positively stabilize the contacts. An investigation of various detent arrangements demonstrated that the common type of series coil detent is ineffective in stabilizing inductively loaded contacts unless combined with a resistance-capacity contact protection circuit. Parallel detent coils provide greater efficiency but are generally impractical because of the extra terminals, more turns, and resistor required. Mechanical or permanent-magnet detent has the advantage of being independent of the circuit characteristics but is difficult to control in small amounts.

Contact oscillation frequencies of from 60 to 3,000 cycles per second have been observed with three different types of sensitive contact devices operating with and without contact protection, at currents of from 5 microamperes to 200 milliamperes, at d-c voltages of less than 1 volt to 200 volts, with inductive or resistive loads, and in air or a vacuum.

References

1. ELECTRIC CONTACTS (book), **Ragnar Holm.** Almquist and Wilksells, Stockholm, Sweden, 1946.

2. SENSITIVE RELAY CONTACT PROTECTION SYSTEMS, **J. P. Dallas, T. H. McCully.** *AIEE Transactions*, volume 67, part II, 1948, pages 1204–07.

3. A POLARIZED RELAY AS AN AIRCRAFT CONTROL ELEMENT, **Royce E. Johnson, Frank A. Glassow.** *AIEE Transactions*, volume 67, part II, 1948, pages 1249–55.

4. SENSITIVE RELAYS IN PROCESS CONTROL, **R. T. Pierce.** *Electrical Engineering*, volume 70, number 7, July 1951, pages 625–26.

5. ELECTRICAL CONTACTS (book), **G. Windred.** MacMillan and Company, London, England, 1940.

No Discussion

Piecewise Linear Servomechanisms

JAMES W. SCHWARTZ
STUDENT MEMBER AIEE

LINEAR servomechanisms have found widespread application primarily because of their inherent simplicity. This simplicity is reflected both in the physical structure of an approximately linear system and in its analysis. It is well known that the performance of a nonlinear servomechanism may be considerably better than that of a comparable linear one. Second-order systems may be investigated semianalytically by the phase plane approach. Unfortunately however, analytical methods for the design of more general nonlinear servo systems are virtually nonexistent. An analytical method based on a rather restrictive approximation was proposed by Kochenburger[1] for analyzing contactor servos. Extension of Kochenburger's method to more general problems is discussed by Johnson.[2]

Transient oscillations present in linear systems are in general due directly to energy storage elements such as inertias and inductances. The net momentum transferred to an inertia due to a linear restoring term actually tends to sustain the oscillation. The only deterrent is the dissipative or damping action of frictional terms. Furthermore, there exists an upper limit (critical damping) to the amount of damping that can be used effectively to improve transient response. Other considerations usually demand the use of somewhat less than critical damping.[3]

The use of nonlinear techniques offers the possibility of improvement of virtually every performance characteristic including faster response with complete elimination of transient oscillation. Several nonlinear schemes have been discussed in the literature.[4–6] Perhaps the most interesting of these are of the anticipator types where discontinuous control action is used. Because the differential equations characterizing such systems are in general insoluble, the discussions have been qualitative. Usually only second-order systems are considered so that phase plane methods can be used.

This paper discusses the improvement of a linear servomechanism by the introduction of piecewise linear control, that is, discontinuous anticipator action where the system equations are linear—but

---

Paper 52-344, recommended by the AIEE Feedback Control System Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Middle Eastern District Meeting, Toledo, Ohio, October 28–30, 1952. Manuscript submitted July 16, 1952; made available for printing September 5, 1952.

JAMES W. SCHWARTZ is with RCA Laboratories, Radio Corporation of America, Princeton, N. J.

The author wishes to express his sincere appreciation to Prof. W. E. Meserve of Cornell University, for his encouragement and guidance in the preparation of this paper.

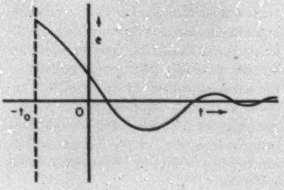


Figure 1. Ordinary transient in moderately damped second-order system

different—before and after the discontinuity. It is pointed out that such equations can be solved. A simple second-order system is analyzed in detail for the interesting case of discontinuous insertion of a constant or coulomb friction term. Experimental results substantiate the analysis.

Anticipator Servomechanisms

Variables such as $c(t)$ and $(dc/dt)(t)$, the controlled variable and its time derivative, $e(t)$ and $(de/dt)(t)$, the actuating signal and its derivative, the error and error rate, may be considered as parameters defining the state of a servomechanism at time t. The number of such parameters necessary to completely specify the state of a system depends on the order of the differential equation associated with it. An anticipator servomechanism monitors two or more such variables, makes predictions as to the state of the system at subsequent times, and sends control signals to the servomotor (or its amplifier) based on the prediction. The type of anticipator considered in this paper is one that monitors the error and error rate and when a transient overshoot appears imminent, transmits a corrective signal to the servomotor, thereby preventing the overshoot.

Piecewise Linear Equations

The state of any linear system with one dynamic degree of freedom can be specified completely by the operational equation

$$B_1e(t)=Fv(t)B_1 \tag{1}$$

together with m boundary (usually initial) conditions. B_1 and F are arbitrary linear operators, m is the order of equation 1, and $v(t)$ is the command signal.

The transient response is obtained from the reduced equation where v is set equal to zero.

$$B_1e=0 \tag{2}$$

If sometime during the course of the transient the system is abruptly altered by anticipator action, the system equations subsequently take the form

$$B_2e(t)=A(t) \tag{3}$$

Here e is still interpreted as v-c. B_2 is an operator which may or may not be the same as B_1. $A(t)$ is an arbitrary function characterizing the anticipator action. If equation 3 is linear, the system is a piecewise linear system. Under such circumstances, equation 3 always can be solved. The solution will contain constants which are determined from the boundary conditions of the problem.

For any real system, two of the boundary equations express the continuity of e and its first derivative during anticipation. In general there will be no continuity of the higher derivatives at the instant of anticipation. The remaining boundary conditions depend upon the type of anticipator action. One or more of these conditions will always exist. Conversely, the arbitrary choice of such boundary conditions determines the character of the anticipator action.

The procedure will be clarified by the investigation of a second-order system which has one arbitrary boundary condition but with several interesting ways of choosing that condition.

Second-Order Anticipator System

For a simple second-order linear system, the complete differential equation corresponding to equation 1 is

$$J\frac{d^2e}{dt^2}+f\frac{de}{dt}+ke=J\frac{d^2v}{dt^2}+f\frac{dv}{dt} \tag{4}$$

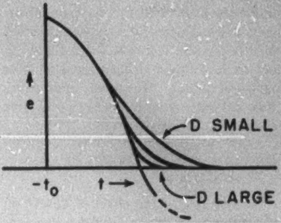


Figure 3. Transient response showing the variation in transient shape with changes in the parameter D

The reduced equation is

$$J\frac{d^2e}{dt^2}+f\frac{de}{dt}+ke=0 \tag{5}$$

The symbols J, f, and k characterize respectively the inertia, viscous friction, and spring constant of the system.

It is convenient to express equation 5 in terms of the damping ratio ζ, natural frequency ω_n, and so forth. The following symbols are defined for this purpose

$$\zeta=\frac{f}{2\sqrt{Jk}} \tag{6}$$

$$\omega_n=\sqrt{k/J} \tag{7}$$

$$\eta=\sqrt{1-\zeta^2} \tag{8}$$

$$\omega_d=\zeta\omega_n \tag{9}$$

$$\omega=\eta\omega_n \tag{10}$$

Equation 5 now may be written as

$$\frac{d^2e}{dt^2}+2\omega_{d(de/dt)}+\omega_n^2\epsilon=0 \tag{11}$$

Assuming that <1 and that a step positional input of amplitude C is introduced at time $t=-t_0$, the solution of equation 11 is

$$e=A\epsilon^{-\omega_d t}\sin(\omega t+\phi) \tag{12}$$

$$A=C\epsilon^{-\omega_d t_0} \tag{13}$$

$$\phi=\omega t_0+\tan^{-1}\left(\frac{\eta}{\zeta}\right) \tag{14}$$

A graph of equation 12 has the familiar form shown in Figure 1. To simplify subsequent equations, the zero time axis

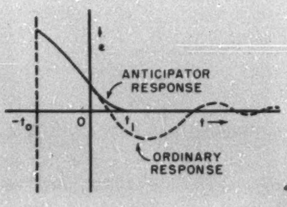


Figure 2 (left). Comparison of anticipator and ordinary system transient

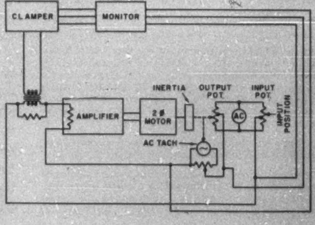
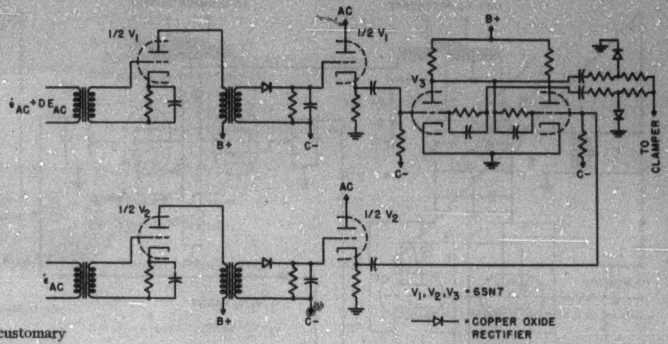


Figure 4 (right). Experimental second-order anticipator system

Figure 5. Monitor circuit



has been shifted from its customary position by t_0 units.

If a constant retarding torque is superimposed (added to the already existing torques) on the system in the time interval $0 \leq t \leq t_1$ is

$$J\frac{d^2e}{dt^2}+f\frac{de}{dt}+ke=JN\omega^2n \tag{15}$$

or

$$\frac{d^2\epsilon}{dt^2}+2\omega_d\frac{d\epsilon}{dt}+\omega_n^2\epsilon=N\omega_n^2 \tag{16}$$

$JN\omega_n^2$ is the magnitude of the superimposed constant torque. The solution to equation 16 is

$$e=N-\alpha\epsilon^{-\omega_d t}\sin(\omega t+\phi) \tag{17}$$

Here α, ϕ, and N are constants to be specified from the boundary conditions. Two of the boundary conditions express the continuity of $e(\mathrm{o})$ and $e(\mathrm{o})$, the error and error rate at the instant of operation of the anticipator.

It is shown in Appendix I that the anticipator must function in such a way that

$$e(\mathrm{o})-D_1\frac{de}{dt}(\mathrm{o})=N \tag{18}$$

if the transient is to have the general shape shown in Figure 2. Equation 18 may be called the characteristic equation of the anticipator.

Under certain circumstances D_1 is a constant. This occurs when the magnitude of the superimposed constant torque is made proportional to $e(\mathrm{o})$, the value of $e(t)$ at the instant of anticipator operation. But if N is proportional to $e(\mathrm{o})$, then by equation 18, it is also proportional to $(de/dt)(\mathrm{o})$. Equation 18 then takes the very simple form

$$\frac{de}{dt}(\mathrm{o})+De(\mathrm{o})=0 \tag{19}$$

D is a constant which is independent of the state of the system. The magnitude of D determines the shape of the transient. This is shown in Figure 3.

An anticipator servomechanism of this type simply monitors the signal

$$M(t)=\frac{de}{dt}(t)+De(t) \tag{20}$$

Whenever $M(t)$ goes through zero, the anticipator sends to the servomotor a constant signal proportional to the error (or equivalently the error rate) at that instant. The signal is removed when $(de/dt)(t)=0$.

Anticipator Characteristic Equation

The extremely simple form of the anticipator characteristic equation 19 is no accident. Because the amplitude of the superimposed torque is made proportional to $e(\mathrm{o})$, the system equation 16 is independent of scale. That is, all transients take the same time and have the same shape regardless of the size of the step input. Only linear or piecewise linear systems have this property.

The simplicity of equation 19 is by no means limited to second-order systems. Similar linear characteristic equations exist for any order piecewise linear system. It is only necessary that the amplitude of the superimposed term or terms be proportional to the appropriate derivatives of e at $t=0$ so that the system equations will be independent of scale. Noise, circuit complexity, and so forth, in practice, set an upper limit to the number of derivatives one can effectively use in a system.

Other Characteristic Equations for the Second-Order System

The superimposed torque signal in the second-order system was assumed to be made proportional to $e(\mathrm{o})$. This requires the use of a clamper circuit[7] or similar device. If, instead, the magnitude of the superimposed signal is always the same, the clamper unit is not needed. In this case, however, D_1 in equation 18 is no longer constant and equation 19 is no longer the correct chracteristic equation. In fact, under these circumstances no simple characteristic equation exists for the general second-order system.

Since the anticipator prevents overshoots, very little viscous damping need be present. A rather simple asymptotic characteristic equation (exact in the limit of no viscous damping) results when the viscous terms can be neglected. In Appendix II this is shown to be

$$\frac{\frac{de^2}{dt}(\mathrm{o})}{\omega^2}-2e(\mathrm{o})N+e(\mathrm{o})^2=0 \tag{21}$$

The monitored signal in this case is

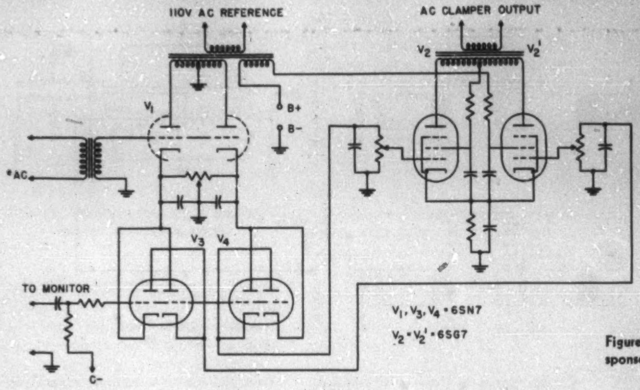
$$s(t)=\left[\frac{\frac{de}{dt}(t)}{\omega}\right]^2+[e(t)]^2-2e(t)N \tag{22}$$

Whenever $s(t)$ passes through zero, the constant torque signal (of the correct polarity, but always of the same magnitude) is inserted. When $(de/dt)(t)$ is zero, it is removed.

Although this type of operation eliminates the need for the clamper circuit, two squaring networks are needed.

Some Additional Considerations

Although a step input was assumed, it is obvious from the derivation in Appendix I that the operation of the anticipator in no way depends upon this assumption. The characteristic equation defines a surface in the phase space composed of all points through which the state of the system can pass. For the second-order



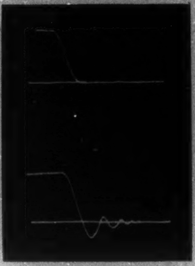


Figure 6 (left). Clamper circuit

Figure 7 (above). Comparison of transient response of the experimental model with and without anticipator action

system, the phase space reduces to a phase plane. The method by which the state of the system arrives at this surface is, of course, of no importance to the subsequent motion.

The great advantage of a piecewise linear system where the scale has been eliminated is the fact that the characteristic equation defines a plane in phase space. For a second-order system, this is a straight line in the phase plane. Consequently, the monitored signal is merely a linear combination of the signals defining the system in phase space.

Many systems have significant amounts of coulomb friction present at all times. This is one reason why it is normally desirable to allow the system to oscillate slightly during a transient. Natural coulomb friction can be compensated for exactly by merely subtracting its value from the synthetic coulomb friction (superimposed anticipator signal) applied by the anticipator. This is simply an adjustment and requires no circuit changes.

Although many systems are approximately linear within some range they very often have saturation or limiting torques and/or speeds. It can be shown that the characteristic surface in phase space is always within or coincides with the surface defined by the limiting torques, speeds, and so forth for all normally accessible points. Consequently, saturation effect will not prevent the anticipator from functioning properly.

Experimental Investigation

The results of the piecewise linear servomechanism theory were applied in the construction of a simple second-order experimental anticipator system. Figure 4 shows the basic diagram of the system. A-c components were used throughout.

The output position signal is obtained from the center tap of an a-c excited potentiometer. Actually, a dual potentiometer was used with a-c excitation on one and d-c excitation on the other. The latter was used to supply the Y axis deflection voltage on an oscilloscope. This eliminated the need for a phase sensitive detector in the oscilloscope setup used to observe the transient behavior.

Figure 5 is a diagram of the monitor circuit. The monitored d-c signal $s=(de/dt)+de$, is amplified by the first section of V_1 and then rectified by a copper oxide rectifier. The second half of V_1 is adjustable so that it is normally just above cutoff voltage. Unless s is very close to zero, the rectified signal drives the second half of V_2 beyond cutoff. Alternating current is supplied to the plate so that when $s=0$ the first section of V_2, a single shot multivibrator, receives a positive pulse causing it to conduct. A sharp positive pulse is generated and sent to the clamper circuit. V_2 functions in exactly the same manner as V_1, causing a positive pulse to appear at the clamper input when $de/dt=0$. A filter at the multivibrator eliminates all negative pulses from the monitor output.

Figure 6 shows the clamper circuit used. V_1 is a phase sensitive detector which produces a d-c output proportional to $e(t)$. V_2 and V_2' form a variable μ pentode modulator whose output is proportional to the stored change on its grid capacitors. The grid capacitors are connected momentarily to V_1 by means of switch tubes V_3 and V_4 which are closed by pulses from the monitor circuit.

The system appeared to perform normally but with much better transient response. Figure 7 shows the transient response of the system following a step positional input. It illustrates the performance of the system with and without anticipator action.

Other Piecewise Linear Servomechanisms

The general theory of piecewise linear system equations is by no means limited to anticipator systems designed solely to improve transient performance. Several other types of control might prove useful. For example, linear systems with negative inertia and/or damping feedback signals have very interesting properties. With ordinary linear operation, however, such systems often have faults such as slow response or various instabilities. By appropriate anticipator action, the negative signals could be inserted and removed automatically in such a way as to eliminate such faults and still retain the benefits of the scheme.

Summary

Although the performance of a piecewise linear anticipator system is not necessarily any better than that of a more general nonlinear anticipator system, it can be significantly better than that of a comparable linear system. Furthermore, no analytical method exists for the investigation of the general nonlinear system and qualitative phase plane studies are limited to second-order systems. For a piecewise linear system of

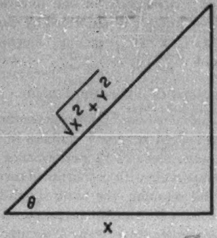


Figure 8. Relation of angle θ to X and Y

any order, complete analytical investigation is possible.

Since the scale can be eliminated in piecewise linear as well as in linear system equations, a special simplicity exists in the design and operation of such systems.

Appendix I. Second-Order Characteristic Equation

The solution to the system equation in the region $0 \leq t \leq t_1$

$$(d^2e/dt^2)_1+2\omega_d(de/dt)_1+\omega_n^2e_1=N\omega_n^2 \tag{23}$$

has the general form

$$e_1(t)=N-\alpha\epsilon^{-\omega_d t}\sin(\omega t+\theta) \tag{24}$$

For reasons of continuity, however, $e_1(t)$ must be chosen in such a way that

$$e_1(o)=e(o) \tag{25}$$

$$(de/dt)_1(o)=(de/dt)(o) \tag{26}$$

Here $e(o)$ and $de/dt(o)$ are the actuating signal and its derivative immediately before the anticipator operates.
From equation 23

$$(de/dt)_1(t)=\omega_d\alpha\epsilon^{-\omega_d t}\sin(\omega t+\theta)- \omega\alpha\epsilon^{-\omega_d t}\cos(\omega t+\theta) \tag{27}$$

Let

$$x=\alpha\cos(\theta) \tag{28}$$

$$y=\alpha\sin(\theta) \tag{29}$$

Then

$$e_1(o)=N-y=e(o) \tag{30}$$

$$(de/dt)_1(o)=\omega_d y-\omega x=(de/dt)(o) \tag{31}$$

or

$$x=\frac{\omega_d[N-e(o)]-(de/dt)(o)}{\omega} \tag{32}$$

$$y=N-e(o) \tag{33}$$

But from equations 28 and 29

$$x^2+y^2=\alpha^2(\sin^2\theta+\cos^2\theta)=\alpha^2 \tag{34}$$

$$\alpha=\sqrt{x^2+y^2} \tag{35}$$

See Figure 8.

For the transient to have the form shown in Figure 2 it is necessary that

$$(de/dt)_1(t_1)=e(t_1)=0 \tag{36}$$

Hence we have the two additional equations

$$N-\alpha\epsilon^{-\omega_d t}1\sin(\omega t_1+\theta)=0 \tag{37}$$

$$\omega_d\sin(\omega t_1+\theta)-\omega\cos(\omega t_1+\theta)=0 \tag{38}$$

From equation 38 then

$$\frac{\omega}{\omega_d}=\tan(\omega t_1+\theta)=\text{constant}=\tan\beta \tag{39}$$

Hence

$$\beta=\omega t_1+\theta=\text{constant} \tag{40}$$

But from equation 37

$$N-\epsilon-\frac{\beta-\theta}{\omega/\omega d}\alpha\sin\beta=0 \tag{41}$$

One way of satisfying this equation is to make $N=\gamma\alpha$ where γ is a constant. But from equation 41 this requires that

$$\gamma=\sin\beta\,\epsilon-\frac{\beta-\theta}{\omega/\omega d}=\text{constant} \tag{42}$$

which is satisfied only if θ = constant.

Under these circumstances, however, it is seen from equation 24 that all transients will have the same shape regardless of amplitude.

From equations 32 and 33

$$ctn\,\theta=\frac{x}{y}=\frac{\omega_d[N-e(o)]-(de/dt)(o)}{\omega(N-e(o))} \tag{43}$$

$$ctn\,\theta=\frac{\omega_d}{\omega}\left[\frac{-(de/dt)(o)}{N-e(o)}+1\right] \tag{44}$$

If θ is a constant we have

$$\frac{(de/dt)(o)}{N-e(o)}=-\frac{1}{D_1}=\text{constant} \tag{45}$$

$$e(o)-D_1(de/dt)(o)=N \tag{46}$$

This is the characteristic equation of the anticipator.

Appendix II. Asymptotic Second-Order Equation

If θ is not constant, there is no simple characteristic equation. The simplicity of equation 46 is related to the elimination of scale by equation 42. For the case where viscous damping is small, an asymptotic equation exists which is not too involved.

The solution equation 37 then takes the form

$$e_1(t)=N[1-\sin(\omega t+\theta)] \tag{47}$$

$$(de/dt)(t)=-\omega N\cos(\omega t+\theta) \tag{48}$$

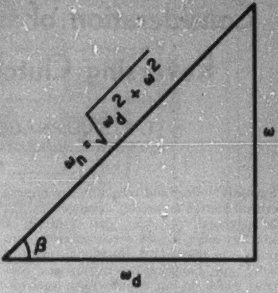


Figure 9. Relation of angle β to ω_d and ω

When $t=t_1$ it is seen that

$$e_1(t_1)=N(1-1)=0 \tag{49}$$

$$(de/dt)_1(t_1)=-\omega N\cdot 0=0 \tag{50}$$

It is required, however, that

$$e_1(o)=e(o)=N[1-\sin\theta] \tag{51}$$

and

$$e_1(o)=e(o)=-\omega N\cos\theta \tag{52}$$

Consequently eliminating $\sin\theta$ and $\cos\theta$

$$e(o)=N\left[1-\sqrt{1-\left(\frac{e(o)}{N\omega}\right)^2}\right] \tag{53}$$

$$\frac{e(o)^2}{N}-\frac{2e(o)}{N}+1=1-\frac{(de/dt)(o)^2}{N\omega} \tag{54}$$

$$e(o)^2-2e(o)N+\frac{(de/dt)(o)^2}{\omega^2}=0 \tag{55}$$

This is the asymptotic anticipator characteristic equation.

References

1. A FREQUENCY-RESPONSE METHOD FOR ANALYZING AND SYNTHESIZING CONTACTOR SERVOMECHANISMS, **R. J. Kochenburger.** *AIEE Transactions*, volume 69, part I, 1950, pages 270–83.

2. SINUSOIDAL ANALYSIS OF FEEDBACK-CONTROL SYSTEMS CONTAINING NONLINEAR ELEMENTS, **E. C. Johnson.** *AIEE Transactions*, volume 71, part II, 1952, pages 169–81.

3. PRINCIPLES OF SERVOMECHANISMS (book), **G. S. Brown, D. P. Campbell.** John Wiley and Sons, Inc., New York, N. Y., 1948.

4. NONLINEAR TECHNIQUES FOR IMPROVING SERVO PERFORMANCE, **Donald McDonald.** Cook Research Laboratories (Chicago, Ill.), 1950.

5. MULTIPLE MODE OPERATION OF SERVOMECHANISMS, **Donald McDonald.** Cook Research Laboratories (Chicago, Ill.), 1951.

6. A PHASE-PLANE APPROACH TO THE COMPENSATION OF SATURATING SERVOMECHANISMS, **Arthur M. Hopkin.** *AIEE Transactions*, volume 70, part I, 1951, pages 631–39.

7. ELECTRON TUBE CIRCUITS (book), **Samuel Seely.** McGraw-Hill Book Company, Inc., New York, N. Y., 1950.

8. INTRODUCTION TO NON-LINEAR MECHANICS (book), **N. Minorsky.** J. W. Edwards, Inc., Ann Arbor, Michigan, 1947.

No Discussion

Consideration of Off-On-Modulated Reversing Clutch Servo Systems

T. R. STUELPNAGEL
ASSOCIATE MEMBER AIEE

J. P. DALLAS
MEMBER AIEE

Synopsis: Cost and reliability are present critical factors in aircraft development. It is suggested, therefore, that a study of simpler, more rugged servo controls is in order. In this regard, the improvement of older-type, nonlinear electromagnetic controls which do not depend upon vacuum-tube equipment should be considered. Off-on-modulated servo clutch controls of this type have substantial advantages in weight, simplicity, and acceleration characteristics. They may be made to approximate closely the performance of a proportional type of control by the use of velocity feedback. Power gains of up to 50 times in the magnetic servo clutch allow direct operation from the signal sensing unit.

Introduction

Application Background

The application of servo mechanisms to aircraft control is largely the result of the last 10 years of aircraft development. Ten years ago many multiple-engine airplanes had no servo systems. In this period, each year has seen an approximate doubling of the planned use of such controls. Today a major portion of the total military airplane cost is invested in servomechanisms of one type or another. The industry is progressing toward an immediate planned future where the pilot will exercise only a monitoring function over the servomechanism which will take off, pilot, and land many aircraft.[1,2]

Mechanical Override

It was necessary to supplement early electric and hydraulic servomechanisms by manually operated mechanical controls. It is doubtful if today's servomechanisms, with their increased complication, have any greater net reliability. Yet, in many cases, the requirement for parallel auxiliary mechanical control has been eliminated.[1] While this elimination has been unavoidable sometimes, due to the complicated functions required of present controls, it usually cannot be justified on the basis of improved reliability of present servo controls. Unfortunately, as the complexity of a control problem increases, a point will be reached where mechanical control is no longer the most reliable. A further increase in requirements can make direct mechanical controls impractical. The control of modern jet engines is an example.

Critical Problem: Cost and Reliability

The rapid expansion of aircraft servomechanism application has caused air force operation and financial problems of a critical nature. The size and effectiveness of the United States Air Force is being limited by the rapid rise in cost of present military aircraft. No small part of this rise in cost is due to the high cost of complicated electronic servo devices. Also, the general failure of electric servo reliability to progress at a rate commensurate with the increased responsibilities now being given aircraft servo control is one of the principal causes of an alarming number of noncombat casualties, noncombat man-hours required for air force operation, and nonoperative airplanes in a typical combat group.[1,2]

Purpose of Article

It will be the purpose of this article to draw attention to an improved type of nonlinear or on-off servomechanism. Development of the on-off servomechanism has been by-passed to a large extent because of the complexity of nonlinear mathematical analysis. Electric contact and mechanical problems also have discouraged this control approach.

Experience with on-off servos, however, has demonstrated that relay and magnetic clutch problems can be solved if they are approached seriously. The resulting device has the possibilities of being less complicated, more rugged and reliable, and of having better acceleration characteristics than proportioned controls of similar application.

Proposed Servomechanism

This article deals with an off-on-modulated type of servomechanism. This off-on servo is distinguished from conventional servos of this type in that it incorporates velocity feed back and high-speed servo reversing clutches. Servo power is obtained from a driving motor operated continuously in one direction. The servo function is performed by direction of rotation reversing magnetic clutches.

Advantages of Proposed Servo

The combined use of velocity feedback and high-speed reversing clutches in an on-off servo elevates its performance to that comparable with proportional servos while still retaining the basic simplicity of the on-off control principle. For example, single-stage control accuracies of 0.1 per cent of total travel can be obtained. This accuracy results because the low time constants in the servo system as a whole provide single step travels as small as 0.05 per cent of total travel.

A further advantage of an off-on-modulated servo is that acceleration performance may be obtained from an all-electric servo which presently can be obtained only from combined electrohydraulic or electropneumatic systems. As an example, electric actuator accelerations of 50,000 to 100,000 radians per second per second can be obtained. By comparison, conventional hydraulic actuators have accelerations of about 20,000 radians per second per second.[3]

Outline of an Off-On-Modulated Reversing Clutch Servo

Signal Source

Referring to the functional block diagram of Figure 1, the signal source for the subject servo may be any of the common a-c or d-c types. Signal generating and follow-up pairs of resistance potentiometers or Selsyn units may be used. Resistance thermometer bridges or variable reluctance bridges yielding displacement, acceleration, or strain signals are suitable signal sources.

Signal Sensing Device

The signal sensing device in its simplest form may be a polarized relay or, for a-c signal systems, a ring demodulator and magnetic amplifier. Because of the power amplifying properties of a magnetic servo clutch, the output of the signal sensing device often may be used to control the servo clutch power unit directly. For example, in its simplest form a polarized relay signal sensing

---

Paper 52-329, recommended by the AIEE Air Transportation Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Middle Eastern District Meeting, Toledo, Ohio, October 28–30, 1952. Manuscript submitted August 6, 1952; made available for printing September 8, 1952.

T. R. Stuelpnagel is at 3922 San Anseline Avenue, Long Beach, Calif., and J. P. Dallas is at 8511 Vicksburg Avenue, Los Angeles, Calif.

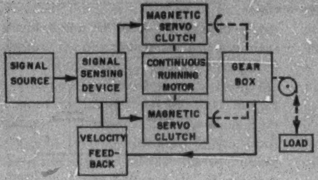


Figure 1 (left). Block diagram of off-on-modulated servomechanism

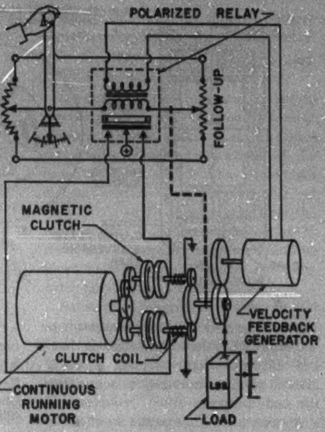


Figure 2 (right). Basic circuit diagram

device may control directly a 300-watt output servo clutch.

Velocity Feedback

The block diagram of Figure 1 also shows a velocity feedback signal generator. This generator provides a velocity proportional negative feedback signal to the sensing device and is driven by the follow-up action of the servo actuator. The velocity feedback signal can be adjusted to a value which will cause a controlled series of off-on modulations of the servo clutch units in a manner to produce any appropriate deceleration of the actuator as it approaches the desired control point. These off-on pulses of the servo clutches may be made small enough so that for all practical purposes the servo output is a uniform motion.

Servo Reversing Clutches and Gearing

The remaining units in the diagram of Figure 1 are the two magnetic reversing clutches, the continuously running motor, and the actuator gearing. These units, together with the feedback generator and the follow-up signal source, are assembled into the servo actuator.

Servo Power Gain

A power gain of over 1,000,000 may be obtained reliably from even a simple form of this system such as the polarized relay, clutched servo described in Figure 2. For example, a typical modern aircraft polarized relay can be made to operate reliably on a 300-microwatt input signal.[4] Such a relay equipped with an adequate contact protection network [5,6] is capable of controlling as much as 10 watts of electric power. This amount of power in turn can control a magnetic servo clutch of several hundred watts. These values are conservative and are not based on speculative development of components such as magnetic powder clutches, which may or may not prove to be superior to the simple disc friction clutch using time-proved materials and design. The point is that these results can be expected from using the nonspeculative design of components to be described in more detail in the following sections.

Magnetic Servo Clutch Power Unit

The Proportional Servo Clutch

Several types of servo clutches have been employed in the past in aircraft controls. Most popular of these in the initial stage of servo clutch development was the proportional clutch. In one proportional clutch design two clutches were geared to drive the output shaft in opposite directions. These clutches were maintained in slight engagement in the normal or zero signal position. An error signal would cause the increased engagement of one oppositely driving clutch and decrease the engagement of the other. This action caused a positive rotating force on the output shaft which corrected the position of the actuator and thus eliminated the error signal. The object of this type of servo clutch was to obtain a torque output proportional to the control signal.

The porportional type of servo clutch has the advantage of providing a theoretically infinite positioning characteristic necessary for system stability in high-accuracy servos. Also, and not the least of its attractiveness, is that it is considerably easier to treat in a mathematical analysis than the nonlinear on-off clutch. The proportional clutch has this disadvantage, however: the attempt to maintain proportional control between mechanical drive members involves slippage, severe clutch heat dissipation problems and low efficiency.

Positively Engaged Dry-Disc Clutch

A more direct approach to the problem of servo clutch design is to admit that the most efficient clutch operation can be obtained with the simple and positive engagement of a dry-disc clutch. Then, if provisions in the servo system are made to off-on-modulate this nonlinear control device, a close resemblance to the desirable proportional control can be obtained. On the whole, pulsing the clutch in and

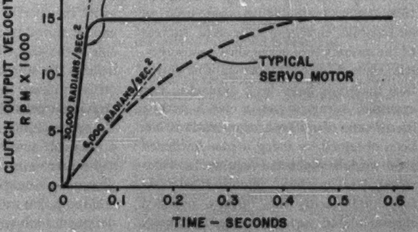


Figure 3. Oscillogram of servo clutch versus servomotor acceleration

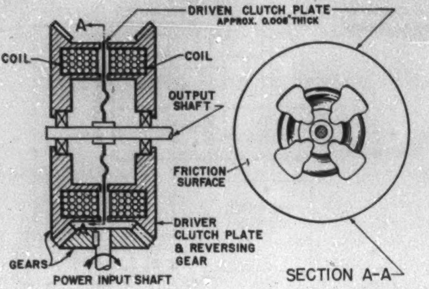


Figure 4 (left). Low-inertia clutch

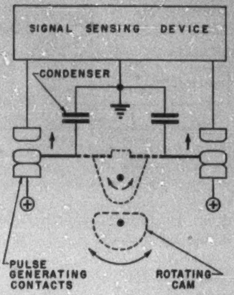


Figure 5 (right). Pulse velocity feedback generator

out for a number of pulses may result in less internal problems and better mechanical efficiency than a clutch which is allowed to slip continously to accomplish the same purpose.

Another advantage of the on-off servo clutch is that its control accuracy is not a function of the actuator gearing load or the variable load that might be coupled to the actuator output. With a positive-action servo clutch the servo control accuracy is determined only by the electrical dead space in the sensing element.

The important concept is that a simple and straightforward clutch can be made very effective for servo control. The simplicity afforded by magnetic dry-disc clutches allows a designer to put most of this effort into the details of design which make a clutch reliable instead of having to put the majority of his effort into making some new type of clutch function.

Dry-Disc Clutch Materials

The primary design problem in a dry-disc servo clutch is controlling the wear that results when the servo load is coupled into the continuously rotating servo motor. The life that may be obtained from a dry-disc clutch is determined by this wear. Clutch failures from excessive wear can result from mechanical failure of the clutch plates or an excessive increase in clutch gap.

In the past, popular clutch materials have been soft metal, cork, or composition material similar to that used in automobile clutches. The real success of the aircraft type servo clutch, however, has resulted because more durable clutch face materials have been applied.[7] For example, a rugged servo clutch with a life of tens of millions of operations has been obtained by using a case hardened steel surface operated against a cold-rolled steel surface. Another successful clutch facing has been a polished chrome-plated surface operating against a cold-rolled steel surface. Clutch combinations of stainless steel against case-hardened steel also have proved successful.

Speed of Response

One of the most important considerations in a servomechanism design is the speed of response. The power controlling devices must start and stop the mechanism rapidly. With such high-speed operation the servo dynamic error is reduced and stability improved. There is also less chance that the actuator will over-shoot and cause mechanical damage at the limits of travel.

High-speed response is difficult to obtain with a typical power relay and servo motor type of control because of the inertia of the motor armature. In typical mechanisms it has been demonstrated that in excess of 90 per cent of the actuator inertia is contained in the motor armature.[8] Furthermore, time lags in the power relays required to control the large currents of the servo motor slow down the operation of the servo. A clutch and brake built into the servo motor may assist by removing the motor armature inertia from the stopping cycle problem but they will not improve the starting characteristic. In any event, the servo motor time requirements must be added to those of the normally required power relay.

When a reversing servo clutch is used with a continuously running motor, however, the inertia of the motor armature can be moved from the debit to the credit side of the design ledger. The servo clutch utilizes the stored energy in the continuously running motor armature for high starting acceleration. Actuation time of the servo clutch itself is no greater than a typical power relay. For example, a d-c servo clutch requiring 10 watts of power for actuation and capable of controlling a 300-watt aircraft motor can be designed to have an operation time of 5 milliseconds and a release time of 5 milliseconds. This would be a better than average performance for commercial aircraft relays designed to handle the same power.

Comparative acceleration characteristics of a servo clutch and servo motor are shown in Figure 3. In this example the servo clutch accelerates at 30,000 radians per second per second or five times faster than the servo motor.

Step Sensitivity of Off-On-Modulated Servo

One of the conflicting design problems in an on-off servomechanism is that of providing adequate power for system acceleration while retaining the small pulse travel needed for stability of the system as a whole. In an on-off servomechanism it is necessary for stability that the minimum pulse travel or step sensitivity fall within the control dead space established by the sensing element. It would seem that this requirement in an on-off servo would be defeated by employing a high-acceleration magnetic servo clutch. Such is not the case, however, if the clutch is designed to provide a few milliseconds of starting slip. With this provision it is possible to obtain a step sensitivity equivalent to or better than a servo motor design, using in both cases the identical motor, actuator, and load characteristics. A starting slip characteristic is shown in the servo-clutch oscillogram of Figure 3.

Clutch Design for Maximum Torque to Inertia Ratio

Where high accelerations of the servo clutch system are desired, excellent torque-to-inertia ratios may be secured with a simple disc clutch as shown in Figure 4. Here a thin disc made from magnetic material is cut away at the center to provide both the necessary spring action and the flexibility to allow the rim of the disc to engage the driving magnets. Clutch disc travel would be of the order of 0.005 inch. The flexing support members are shown beaded in the referenced figure to allow for angular alignment of the disc clutch face with the magnetic driving units. For a servo clutch power unit typical of the size being described in the 200- to 400-watt output range, the clutch disc could be made from material 0.005 to 0.008 inch thick. Such simple designs require careful mechanical refinement but little of the speculative research characteristic of many current experimental proposals and can be expected to have substantial production cost advantages.

EFFICIENCY OF MAGNETIC CLUTCH

The use of magnetic clutches is attractive in servo control because these devices are efficient transducers of power from the electrical to the mechanical form. Servo clutches both transform and amplify power in the same operation. The efficiency of power conversion may be demonstrated by comparing a typical 300 watt output servo motor to a 300-watt output servo clutch. With the servo clutch, 300 watts of mechanical output power may be controlled with 10 watts of electric input power. A 300-watt output from a servo motor, however, would require approximately 500 watts of controlled input power. In this comparison, therefore, the servo clutch alone shows a power amplification of 30 to 1 and an advantage over the motor of 50 to 1 in the amount of electric power required to control a 300-watt mechanical output.

Velocity Eeedback

The application of velocity feedback to off-on-clutched controls elevates performance to a level competitive with proportional controls. Velocity feedback is the brakeman on the servo gear train. Its function is to stop the actuator at the position prescribed by the signal sensing device without overshoot or oscillation. To accomplish this job the velocity feedback first must sense the actuator stopping distance required at any velocity and then signal when to start the stopping operation. In on-off controls velocity feedback provides a damping characteristic by time modulation of the servo clutch. At a point perdetermined by the initial velocity signal the servo clutch is off-on pulsed to obtain the desired deceleration to the selected stopping point.

PULSE FEEDBACK GENERATOR

A simple type of velocity feedback is proposed in Figure 5. With this method a velocity signal is generated by integrating a series of capacitor pulses obtained when electric contacts are operated. These contacts are opened and closed at a rate proportional to servo velocity. Polarity of the feedback signal is obtained by actuating either one of two sets of contacts according to the direction of the driving cam rotation. The feedback generator described is only one of many types that can be employed in servo controls. The most common is the tachometer generator. This type of feedback source generates a signal proportional to velocity and is the type most often employed.

APPLICATION OF FEEDBACK IN OFF-ON SERVOMECHANISMS

One point that should be emphasized about applying feedback in off-on servomechanisms is that such feedback should not be used to compensate for slow electromechanical operation. Every effort should be made to purge the basic servomechanism of unnecessary inertias, compliance, and time lags. Then rate feedback should be added as a frosting on the basic performance. The approach of making the basic system sound before refining it with velocity feedback has three advantages:

1. The feedback circuit, except for insulation, is reduced from a series to a parallel liability. That is, failure of the velocity feedback unit does not cause servomechanism failure.
2. When velocity feedback is finally added the system performance improves. For example, the step sensitivity of the servo may be increased many times.
3. Less feedback voltage is required.

AMOUNT OF FEEDBACK REQUIRED

The amount of velocity feedback required can be determined from an analysis of the servo stopping characteristic. The velocity feedback signal required is merely the error signal cancelled during the stopping interval. If, for example, the servo at a particular velocity cancels error signal at a rate of 0.1 volt per millisecond and the stopping operation consumes an effective time of 50 milliseconds, a feedback signal of 5 volts is required for critical damping.

Signal Sensing Device

The object of this article has been to outline a type of servomechanism which achieves performance without the sacrifice of simplicity. In keeping with this objective a simple clutch power unit and feedback generator has been suggested. It is appropriate, therefore, to complete this discussion with the simplest possible signal sensing device—the polarized relay.

POSSIBILITIES OF POLARIZED RELAY

A simple contact arm and actuating coil comprising a relay is perhaps the most efficient, lightest, and most stable device for getting power amplifications of the order of 50,000. With these advantages it seems worth while to give the polarized relay a chance by providing at least the same degree of engineering attention to its design and application that would be given to competing electronic and magnetic amplifying devices.

Substantial progress in the design and application of polarized relays to aircraft control has been made recently.[4] Much of the current objection to aircraft use of polarized relays may border on prejudice. Such prejudice has resulted from attempts to use former polarized relay devices not adapted to aircraft environment or to serious misapplication. Under the category of misapplication may be included the use of any sensitive contact making device without a thorough investigation of the contact problems and the design of appropriate contact protection networks.[5,6]

It is believed that up-to-date relay units with adequate contact protection networks can be applied to aircraft controls in a manner to insure adequate reliability.

Conclusions

1. The demand for improved reliability in aircraft servomechanism control is increasing at a faster rate than the actual growth of reliability. This reliability factor, together with the cost of aircraft servomechanisms, is a limiting factor in the size and effectiveness of the United States Air Force.
2. Simplification is the first step in improving servo reliability and reducing cost. Electric on-off servo controls are simple and are based on nonspeculative design with rugged electromechanical elements.
3. The combined use of feedback and servo clutches in an on-off servo elevates its performance to that obtainable with continuous control while offering possibilities of lower cost and weight, greater simplicity, and reliability. Single stage control to 0.1 per cent of full travel is practical.
4. All-electric off-on-modulated controls can match response characteristics presently obtainable only with hydraulic or pneumatic controls. Accelerations of 50,000 to 100,000 radians per second per second are obtainable.
5. The off-on-modulated type of servo unit is adaptable to large aircraft servo control applications presently using hydraulic or electrohydraulic servos. The subject unit would offer the advantages of all-electric

control while avoiding the critical motor starting problem, associated line disturbance, and altitude brush problems of a servomotor control.

References

1. REPORT ON ADVISORY STAFF FOR AIRCRAFT ELECTRIC SYSTEMS. Wright Air Development Center and Bureau of Aeronautics, Dayton, Ohio, 1951, pages 21, 45, 56.

2. RESEARCH AND DEVELOPMENT IN THE UNITED STATES AIR FORCE, **Major General Donald L. Putt.** *SAE Quarterly Transactions* (New York, N. Y.), volume 6, number 2, April 1952, pages 304–07.

3. ELECTRONIC INSTRUMENTS (book), **J. A. Greenwood, Jr., J. V. Holdam, Jr., D. MacRae, Jr.** MIT Radiation Laboratory Series Number 21, McGraw-Hill Book Company, Inc., New York, N. Y., 1948, pages 386–436.

4. A POLARIZED RELAY AS AN AIRCRAFT CONTROL ELEMENT, **Royce E. Johnson, Frank A. Glassow.** *AIEE Transactions*, volume 67, part II, 1948, pages 1249–55.

5. SENSITIVE RELAY CONTACT PROTECTION SYSTEMS, **J. P. Dallas, T. H. McCully.** *AIEE Transactions*, volume 67, part II, 1948, pages 1204–07.

6. ELECTROSTATIC OSCILLATION IN SENSITIVE CONTACT MECHANISMS, **Thomas R. Stuelpnagel.** *AIEE Transactions*, volume 71, part II, 1952 (January 1953 Section), pages 397–401.

7. ONE TYPE OF ROTARY MAGNETIC CLUTCH AND ITS ASSOCIATED BRAKE USED ON AIRCRAFT ELECTRIC MOTORS, **Leo Andrews, Fred Shanely.** *Electrical Engineering*, volume 63, number 12, December 1944, pages 893–95.

8. BRAKING DEVICE FOR 400-CYCLE MOTORS, **L. A. Zahorsky.** *Electrical Engineering*, volume 71, number 6, June 1952, pages 506–11.

9. ELECTRONIC AUTO-PILOT CIRCUITS, **W. H. Gille, H. T. Sparrow.** *Electronics* (New York, N. Y.), volume 17, October 1944, page 110.

No Discussion

Impedance Data for 400-Cycle Aircraft Distribution Systems

D. W. EXNER
MEMBER AIEE

G. H. SINGER, JR.
ASSOCIATE MEMBER AIEE

Synopsis: This paper presents impedance data which the authors have compiled for use in designing and testing aircraft electric systems employing 3-phase 400-cycle auxiliary power. Tabulated herein are impedance data for a number of typical configurations employing multiple-wire feeders in 0.5-inch flat spacing and laced 3-phase groups. Sufficient background material and procedure are given to enable the reader to calculate additional data for configurations not actually tabulated.

The positive-sequence data were calculated directly from the geometry of the wire configurations. However, zero-sequence data are dependent upon the particular ground return circuit employed and, therefore, these data are the result of a theoretical calculation modified by an empirically determined skin correction term. The experimental work carried out to obtain these skin correction values is described.

By nature a data presentation such as this is incomplete and somewhat restricted in usefulness. However, it is hoped that those involved in analytical and experimental studies on 3-phase 400-cycle aircraft systems will expand these data to include any new configurations which may be used.

SINCE the advent of large military aircraft with their ever-increasing dependence on auxiliary power for control and tactical functions, there has been a steady increase to higher voltage electric systems. One currently used system employs 120/208-volt 3-phase 400-cycle Y-connected electric power and multiple-wire feeders containing fusible limiters. The problems of fault clearing and limiter co-ordination occurring during short-circuit faults involve the transient characteristics and the sequence impedances of the rotating machinery and the sequence impedances of the circuit elements comprising the distribution system. Similar to the case of industrial power distribution networks, these fault problems are best handled by application of the symmetrical component theory. This study was undertaken because the necessary impedance data were not available for applying symmetrical components to aircraft wire configurations. Reference 1 provides excellent background for the work described here, but it was felt that this material should be expanded to include multiple wires per phase. Also, the effects of ground path on zero-sequence impedance required more thorough investigation, using an actual or a simulated aircraft structure.

Nomenclature

a = radius of solid round conductor, inches
t = time, seconds
L = inductance, henrys
I = current, amperes
μ = permeability = 4×10^{-7} weber per meter per ampere-turn
ln = natural logarithm, base = ϵ = 2.718
log = common logarithms, base = 10
d_s = GMR = self-geometric mean radius of a solid round conductor or a stranded conductor, inches
d^*_s = GMR of a 3-phase group of conductors, inches
GMD = mutual geometric mean distance between conductors i and j, inches
D_{ij} = linear distance between conductors i and j, inches
D = spacing of 3-phase groups, inches
s = spacing of individual conductors in a flat 3-phase group, inches
h = mean height of configuration above skin, inches
R = resistance, ohms per 1,000 feet per phase
R_0 = zero-sequence resistance
R_{0c} = calculated value of zero-sequence resistance assuming perfect ground plane
R_s = skin correction term for zero-sequence resistance
X = reactance, ohms per 1,000 feet per phase
X_0 = zero-sequence reactance
X_{0c} = calculated value of zero-sequence reactance assuming perfect ground place
X_s = skin correction term for zero-sequence reactance
Z = impedance
A = equivalent distance to return circuit

Approach to the Problem

In compiling these data it was assumed that the configurations most likely to be used in 3-phase aircraft systems were laced and that they were 0.5-inch flat spaced groups, as shown in Figure 1. Since various systems are likely to employ from one to five wires per phase at distances from the skin of up to 5 inches, it was felt that data should be available to include all these cases.

To make clear the methods used in determining these data, the pertinent theoretical background should be reviewed briefly. Considering a solid round conductor, it is possible to write an expression for the inductance per unit length as follows

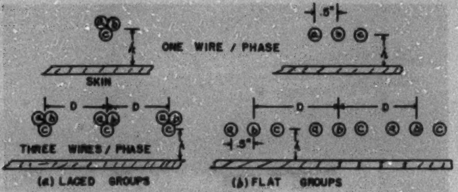
$$L=\frac{\mu}{2\pi}\left(\frac{1}{4}+\ln\frac{A}{a}\right) \qquad (1)$$

The first term gives the inductance due to

Paper 52-322, recommended by the AIEE Air Transportation Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Middle Eastern District Meeting, Toledo, Ohio, October 28–30, 1952. Manuscript submitted July 25, 1952; made available for printing October 6, 1952.

D. W. EXNER and G. H. SINGER, JR., are with the Boeing Airplane Company, Seattle, Wash.

The work on which this paper is based was carried out by the Acoustics-Electrical Unit of the Boeing Airplane Company, Seattle, Wash.



(a) LACED GROUPS (b) FLAT GROUPS

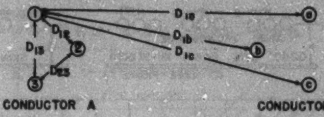


Figure 1 (left)

Figure 3 (above)

the internal flux linkages, and the second term the inductance due to the external linkages. A powerful tool in the calculation of the inductance of a configuration of conductors is the concept of a geometric mean distance. For example, equation 1 can be written

$$L=\frac{\mu}{2\pi}\ln\frac{A}{a\epsilon^{-0.25}}=\frac{\mu}{2\pi}\ln\frac{A}{d_s} \qquad (2)$$

This is equivalent to replacing the solid round conductor by a hollow tube having negligible internal inductance and a radius equal to $a\epsilon^{-0.25}$. This radius is called the self-geometric mean radius of a solid round conductor, abbreviated GMR, or d_s. Equation 2 is important in inductance calculations because it can be used for conductors of any shape or stranding provided the proper GMR is used. The calculation of GMR is merely a geometry problem and is considered in detail in reference 2.

Figure 2 shows a circuit consisting of two arbitrarily shaped current-carrying conductors A and B consisting of n and m individual parts respectively. The following expression for the inductance per unit length for conductor A is developed in reference 3.

$$L=\frac{\mu}{2\pi}\ln\frac{\text{GMD}}{\text{GMR}} \qquad (3)$$

where

$$\text{GMD}=(D_{a1}D_{b1}\ldots D_{a2}D_{b2}\ldots D_{an}D_{bn}\ldots)^{1/mn} \qquad (4)$$

$$\text{GMD}=(d_{s1}D_{12}D_{13}\ldots d_{s2}D_{21}D_{23}\ldots d_{sn}D_{n1}D_{n2}\ldots)^{1/n^2} \qquad (5)$$

The numerator of the logarithmic term is the mutual geometric mean spacing between the n elements of A and the m elements of B, abbreviated GMD. As seen from equation 4, the computation of GMD involves all the mutual spacings between the elements of the conductors A and B, and the GMR involves the geometric mean radii of all the components of A and the spacings between components. The following example illustrates the use of equations 3, 4, and 5 in computing inductance.

Consider the problem of determining the inductance of the single-phase transmission line whose geometry is given in Figure 3. Since the conductors are all similar, equation 5 for GMR reduces to

$$\text{GMR}=[d_s^{n/2}D_{12}\ldots D_{1n}\quad D_{23}\ldots D_{2n}\ldots D_{(n-1)n}]^{2/n^2} \qquad (6)$$

Therefore, for Figure 3

$$\text{GMR}=[d_s^{3/2}D_{12}D_{23}D_{31}]^{2/9} \qquad (7)$$

Using equation 4, GMD becomes

$$\text{GMD}=[D_{1a}D_{1b}D_{1c}\quad D_{2a}D_{2b}D_{2c}\quad D_{3a}D_{3b}D_{3c}]^{1/9} \qquad (8)$$

Substituting equations 7 and 8 in equation 3, the inductance per unit length of the single-phase line of Figure 3 can be calculated if the physical spacing and geometric radii are known.

This method, outlined for a single-phase line, leads directly to the calculation of the zero-sequence impedance of aircraft transmission lines. In large military aircraft the electric power distribution circuits often run normal to the circumferential body stiffeners about 3 to 5 inches from the metal skin or along the main wing spars. Under ground-fault conditions the zero-sequence fault current is, by definition, the single-phase component flowing through the conductors of the faulted circuit into the fault and returning through the grounding system. Consider the line shown in Figure 4 (A) representing the zero-sequence path of an aircraft circuit. To calculate rigorously the impedance as seen from points 1 and 2 of this circuit, a knowledge of the current distribution in the ground plane must be available. Without an extensive mathematical analysis, which even then involves certain approximations, this distribution cannot be defined.

Because of the inability of solving the circuit of Figure 4(A) completely by theoretical methods, the following approach was used. First, the ground plane was considered to be a perfect ground return, and the impedance of the conductor configuration was calculated by image theory. Second, a correction term, determined empirically, was applied to the calculated value to account for the fact that the actual skin is not a perfect ground. A perfect ground plane is defined as a plane sheet of infinite extent having zero resistance and inductance. With such a return circuit, the distribution of the electric and magnetic fields above the ground plane is identical to that existing for a 2-sided line with a return circuit whose configuration is the image of the actual conductors and is located an equal distance below the position of the ground sheet. The image configuration for this line is shown in Figure 4 (B).

Measurements were carried out on 3-phase circuits set up in an *XC-97* air frame to determine the actual 400-cycle zero-sequence impedance for a number of typical configurations. These data then were compared with corresponding values calculated by image theory. Since it was impractical to obtain data for all possible

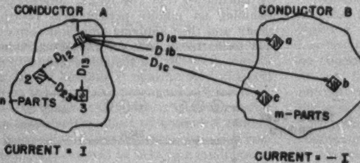


Figure 2

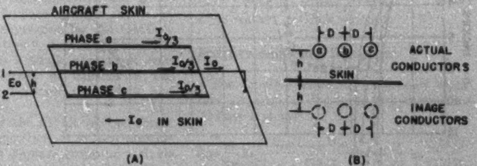
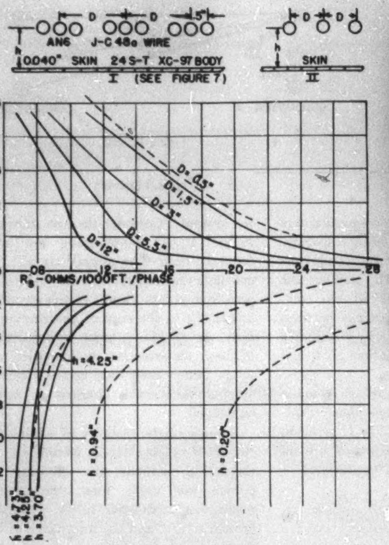


Figure 4



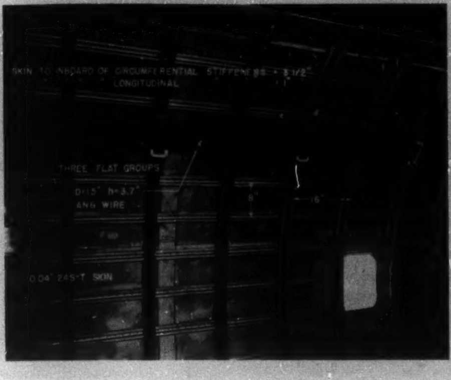


Figure 5 (left). Zero-sequence resistance correction terms from test at 400 cycles per second

Note: Solid curves are for configuration I. Broken curves are for configuration II. Actual test points were within 10 per cent of these curves

Figure 7 (above). Test air-frame structure

spacings and wire configurations, it was necessary to interpolate the data for those cases not actually tested. The skin correction terms are summarized in Figures 5 and 6; the *XC-97* structure is shown in Figure 7. Appendix I illustrates the development of the GMD and GMR equations for the configurations studied, and Appendix II gives the GMD and GMR equations. Appendix III gives examples of the calculation of the positive and zero-sequence data in Table II and in Figures 9 to 25. Appendix IV discusses the test work briefly.

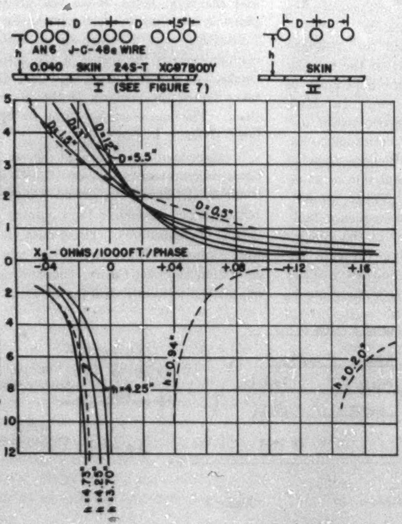


Figure 6 (left). Zero - sequence reactance correction terms from test at 400 cycles per second

Note: Solid curves are for configuration I. Broken curves are for configuration II. Actual test points were within 10 per cent of these curves

Presentation of Data

The 400-cycle impedance data for aircraft wire configurations resulting from this study are summarized in Table II and in Figures 9 to 25. These data should be adequate for analytical and design work on electrical distribution systems for large aircraft. Only data for laced 3-phase groups and 3-phase groups in 0.5-inch flat spacing have been compiled, since these configurations were considered the most likely to be used in present designs. The following statements are made to clarify the use of Table II and Figures 9 to 25:

1. The positive-sequence impedance data are all calculated values based on *AN-J-C-48a* (Army-Navy) wire specifications and on the 400-cycle resistance information contained in reference 4. These data are not affected by the type of body structure or by whether or not the wires are run in nonmagnetic conduit. Also, the elevation of the configuration above the aircraft skin has no effect on the positive-sequence impedance.

2. The zero-sequence data were compiled by calculating the theoretical impedance assuming an infinite perfect skin return circuit and then applying empirical correction terms to account for the actual skin

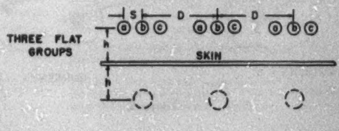
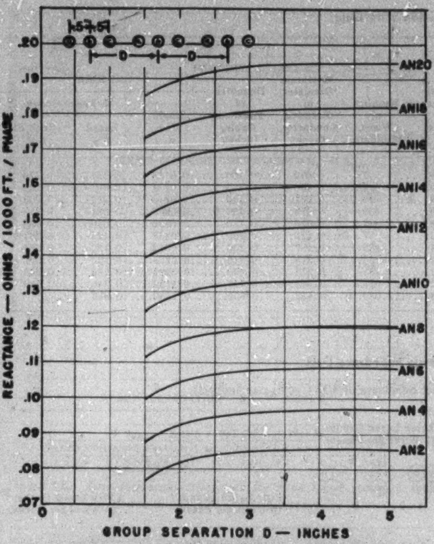
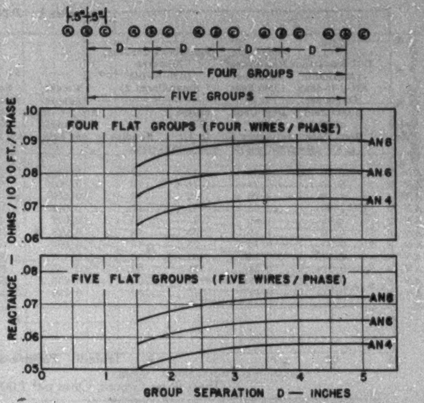


Figure 8 (right)



RESISTANCE — OHMS / 1000 FT. / PHASE									
AN20	AN18	AN16	AN14	AN12	AN10	AN8	AN6	AN4	AN2
3.42	2.15	1.59	.997	.627	.366	.234	.147	.094	.063



	RESISTANCE — OHMS / 1000 FT. / PHASE									
	AN20	AN18	AN16	AN14	AN12	AN10	AN8	AN6	AN4	AN2
FOUR FLAT GROUPS	2.55	1.61	1.19	.747	.470	.275	.175	.110	.071	.047
FIVE FLAT GROUPS	2.05	1.29	.952	.598	.376	.220	.140	.088	.056	.038

Figure 9 (left). Positive-sequence impedance for three flat groups at 400 cycles per second

Figure 10 (above). Positive-sequence impedance for four and five flat groups at 400 cycles per second

structure. These correction terms were determined for a body structure which is considered typical of modern large aircraft designs. Due to the relatively small magnitude of this skin correction term for most wire configurations, and the impracticability of determining it exactly for all variations of body structure, the zero-sequence data presented in the following may be used without qualification.

3. The zero-sequence reactance is read directly from the curves for the proper size, elevation, and group separation. However, the zero-sequence resistance is obtained by adding the value R_{0c} from the table in each figure to the value of R_s from the R_s curve for the proper group separation. R_{0c} is the calculated zero-sequence resistance, assuming a perfect ground return, and R_s is the skin correction term.

4. The positive- and zero-sequence data for configurations employing four and five 3-phase groups were compiled for *AN-8*, *AN-6*, and *AN-4* wire sizes only, since these sizes are most likely to be used whenever the load requirements necessitate the use of four or five wires per phase.

5. Experimentation has shown that the zero-sequence impedance of wires run in grounded conduit is considerably different from the impedance when no conduit is used. At the time of this study, it was not possible to consider the effect of grounded conduit thoroughly. However, since the use of conduit probably will be held to a minimum in aircraft, due to weight considerations, it is felt that the conduit data summarized in Appendix V and Figure 25 are adequate. The positive-sequence impedance is not affected by the presence of nonmagnetic conduit.

6. All resistance data are computed at 20 degrees centigrade.

Appendix I. Development of GMD and GMR Equations

Zero-Sequence Equations

Example 1: Determine GMD and GMR for the 3-wire circuit of Figure 4(B). Using equation 4 the GMD between the three conductors and the image circuit is

$$\mathrm{GMD}=[(2h)^3(4h^2+D^2)^2(4h^2+4D^2)]^{1/9} \qquad (9)$$

From equation 6 the GMR for the circuit of Figure 4(B) can be expressed

$$\mathrm{GMR}=(d_s^{3/2}D2DD)^{2/9}=1.166d_s^{1/3}D^{2/3} \qquad (10)$$

Example 2: Determine GMD and GMR for the three 3-phase groups shown in Figure 8. The expression for the GMD is identical to equation 9, since each group can be considered as a single equivalent conductor for purposes of calculating the mutual spacing between the actual and image conductors.

In calculating the GMR of the actual conductor configuration, it is also possible to consider each 3-phase group as a single equivalent conductor whenever the spacings between the conductors of a group are small compared to the separation of the groups. It has been determined that this approach is valid for all spacings of laced groups and for groups in 0.5-inch flat spacing whenever the group spacing D is at least 3 inches. Therefore, for Figure 8

$$\mathrm{GMR}=1.166d^{*}{}_s^{1/3}D^{2/3} \quad \begin{cases}\text{laced groups: all values } D\\ \text{flat groups: } D>3 \text{ inches}\end{cases} \qquad (11)$$

where $d^*{}_s$ is the GMR of a single group, expressed as

$$d^*{}_s=(d_s^{3/2}D_{ab}D_{bc}D_{ca})^{2/9}$$

For flat groups where D is less than 3 inches equation 12 for GMR is developed, using equation 6 and considering all nine conductors of the configuration shown in Figure 8

$$\mathrm{GMR}=[d_s^{9/2}2^8s^9D^9(D^2-s^2)^6(D^2-4s^2)^2\times(4D^2-s^2)^2]^{2/81} \qquad (12)$$

Equation 12 reduces as follows for 0.5-inch flat groups.

$$\begin{aligned} D&=1.5 \text{ inches:} & \mathrm{GMR}&=1.32d_s^{1/3}\\ D&=2 \text{ inches:} & \mathrm{GMR}&=1.63d_s^{1/3} \qquad (13)\\ D&\geq 3 \text{ inches:} & \mathrm{GMR}&=1.166d_s^{*1/3}D^{2/3}\end{aligned}$$

Positive-sequence Equations

For positive-sequence inductance equation 14 can be developed for transposed 3-phase systems[3]

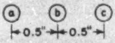
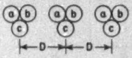
$$L=\frac{\mu}{2\pi}\ln\frac{(D_{mab}D_{mbc}D_{mca})^{1/3}}{\mathrm{GMR}}=\frac{\mu}{2\pi}\ln\frac{\mathrm{GMD}}{\mathrm{GMR}} \qquad (14)$$

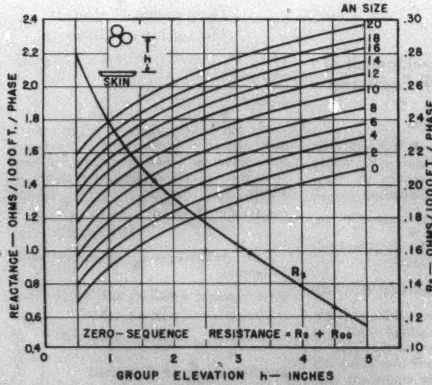
Table I. AN-J-C-48a Wire Data

Cable Size	D-C Resistance (Maximum AN-J-C-48a), Ohms per 1,000 Feet at 20 Deg. Cent.	400-Cycle Resistance, Ohms per 1,000 Feet at 20 Deg. Cent. (per reference 4)	Continuous Loading, Amperes (AN-W-14a Amendment 2)		Weight (Approx. AN-J-C-48a) Pounds per Foot	Nominal Conductor Area, Circular Mils	Number of Wires, Minimum	Maximum Diameter of Stranded Conductor, Inches	Diameter of Finished Cable, Inches	Self-GMR		
			Single Wire in Air	In Bundles						Single Wire d_s	3-Phase Group	
											Laced d^*_s	$^1/_2$-Inch Flat d^*_s
AN-20	10.25	10.25	11	7.5	0.006	1,119	7	0.040	0.100	0.0138	0.0517	0.176
AN-18	6.44	6.44	16	10	0.009	1,779	7	0.050	0.115	0.0174	0.0613	0.190
AN-16	4.76	4.76	22	13	0.012	2,409	19	0.061	0.130	0.0213	0.0711	0.203
AN-14	2.99	2.99	32	17	0.018	3,830	19	0.076	0.150	0.0268	0.0845	0.219
AN-12	1.88	1.88	41	23	0.027	6,088	19	0.096	0.170	0.0338	0.0992	0.238
AN-10	1.10	1.10	55	33	0.043	10,443	37	0.122	0.200	0.0452	0.122	0.261
AN- 8	0.70	0.70	73	46	0.067	16,864	133	0.167	0.255	0.0582	0.156	0.284
AN- 6	0.436	0.440	101	60	0.100	26,813	133	0.218	0.310	0.0734	0.192	0.307
AN- 4	0.274	0.282	135	80	0.155	42,613	133	0.272	0.370	0.0925	0.233	0.332
AN- 2	0.179	0.188	181	100	0.250	66,832	663	0.345	0.445	0.1159	0.284	0.358
AN- 0	0.114	0.127	245	150	0.385	104,118	1,033	0.432	0.550	0.1450	0.353	

Table II. Positive-Sequence Impedance Data

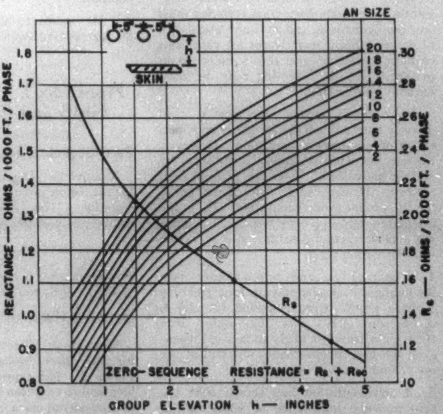
(Three-Phase Groups, Ohms per 1,000 Feet per Phase at 400 Cycles per Second)

	One Flat Group  	One Laced Group 	Three Laced Groups (3 Wires per Phase)  	Four Laced Groups (4 Wires per Phase)	Five Laced Groups (5 Wires per Phase)
AN-20	10.25 $+j0.585$	10.25 $+j0.304$	3.42 $+j0.101$	2.56 $+j0.076$	2.05 $+j0.061$
AN-18	6.44 $+j0.550$	6.44 $+j0.289$	2.15 $+j0.096$	1.61 $+j0.072$	1.29 $+j0.058$
AN-16	4.76 $+j0.519$	4.76 $+j0.278$	1.59 $+j0.093$	1.19 $+j0.070$	0.952 $+j0.056$
AN-14	2.99 $+j0.484$	2.99 $+j0.264$	0.997 $+j0.088$	0.747 $+j0.066$	0.598 $+j0.053$
AN-12	1.88 $+j0.448$	1.88 $+j0.247$	0.627 $+j0.082$	0.470 $+j0.062$	0.376 $+j0.049$
AN-10	1.10 $+j0.405$	1.10 $+j0.228$	0.366 $+j0.076$	0.275 $+j0.057$	0.220 $+j0.046$
AN- 8	0.700 $+j0.366$	0.700 $+j0.226$	0.234 $+j0.075$	0.175 $+j0.057$	0.140 $+j0.045$
AN- 6	0.440 $+j0.329$	0.440 $+j0.221$	0.147 $+j0.074$	0.110 $+j0.055$	0.088 $+j0.044$
AN- 4	0.282 $+j0.294$	0.282 $+j0.212$	0.094 $+j0.071$	0.071 $+j0.053$	0.056 $+j0.042$
AN- 2	0.188 $+j0.260$	0.188 $+j0.206$	0.063 $+j0.069$	0.047 $+j0.052$	0.038 $+j0.041$
AN- 0		0.127 $+j0.204$	0.042 $+j0.068$	0.032 $+j0.051$	0.025 $+j0.041$



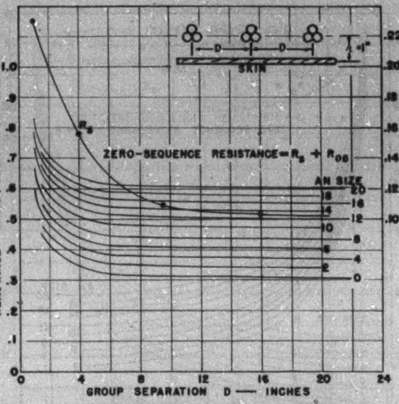
Rec — OHMS/ 1000 FT. / PHASE										
AN20	AN18	AN16	AN14	AN12	AN10	AN 8	AN 6	AN 4	AN 2	AN 0
10.25	6.44	4.76	2.99	1.88	1.10	.700	.440	.282	.188	.127

Figure 11. Zero-sequence impedance for one laced group at 400 cycles per second



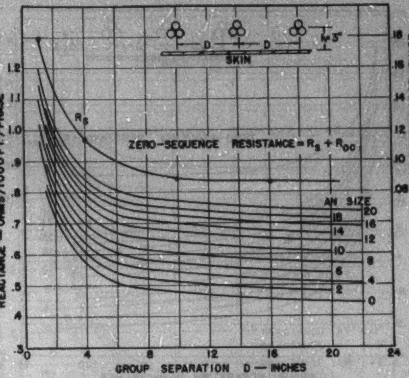
Rec — OHMS / 1000 FT. / PHASE										
AN20	AN18	AN16	AN14	AN12	AN10	AN 8	AN 6	AN 4	AN 2	AN 0
10.25	6.44	4.76	2.99	1.88	1.10	.700	.440	.282	.188	—

Figure 12. Zero-sequence impedance for one flat group at 400 cycles per second



R_{DC} — OHMS / 1000 FT. / PHASE										
AN20	AN18	AN16	AN14	AN12	AN10	AN 8	AN 6	AN 4	AN 2	AN 0
3.42	2.15	1.59	.997	.627	.366	.234	.147	.094	.063	.042

Figure 13. Zero-sequence impedance for three laced groups 1 inch above skin at 400 cycles per second



R_{DC} — OHMS / 1000 FT. / PHASE										
AN20	AN18	AN16	AN14	AN12	AN10	AN 8	AN 6	AN 4	AN 2	AN 0
3.42	2.15	1.59	.997	.627	.366	.234	.147	.094	.063	.042

Figure 15. Zero-sequence impedance for three laced groups 3 inches above skin at 400 cycles per second

where D_{mab}, D_{mbc}, and D_{mca} are the GMD's between the respective phases as indicated by the subscripts, and the denominator is the GMR of one phase configuration. The computation of GMD and GMR for use in equation 14 as follows.

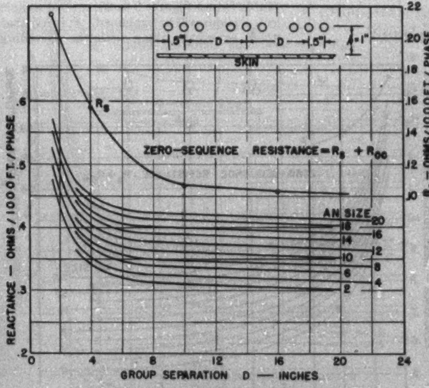
Example 3: Determine GMD and GMR for calculating the positive-sequence inductance of the circuit illustrated in Figure 8. Using equations 14 and 4 the GMD becomes

$$\text{GMD} = \begin{bmatrix} [s(D+s)(2D+s)(D-s)s\times \\ \quad (D+s)(2D-s)(D-s)s]^{1/9} \\ [s(D+s)(2D+s)(D-s)s\times \\ \quad (D+s)(2D-s)(D-s)s]^{1/9} \\ [2s(D-2s)(2D-2s)(D+2s)\times \\ \quad 2s(D-2s)(2D+2s)(D+2s)2s]^{1/9} \end{bmatrix}^{1/3}$$
$$= [2^5s^9(D^2-s^2)^5(4D^2-s^2)^2\times (D^2-4s^2)^2]^{1/27} \quad (15)$$

By considering the three conductors of one phase and using equation 6, the GMR of a phase is

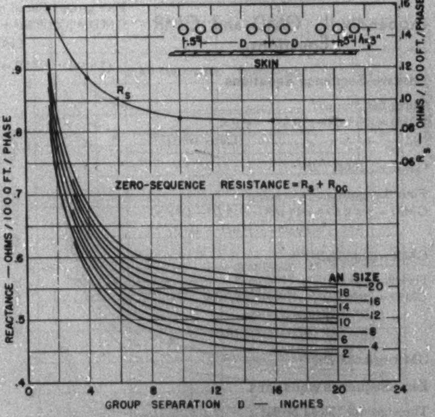
$$\text{GMD} = (d_s^{3/2}D2DD)^{2/9} = 1.166d_s^{1/3}D^{2/3} \quad (16)$$

The positive-sequence inductance of configurations employing laced groups does not depend on the group separation, even for the closet spacings. Therefore, the per-phase impedance of a system employing n wires per phase is found by dividing the per-phase impedance of one laced group by n. This



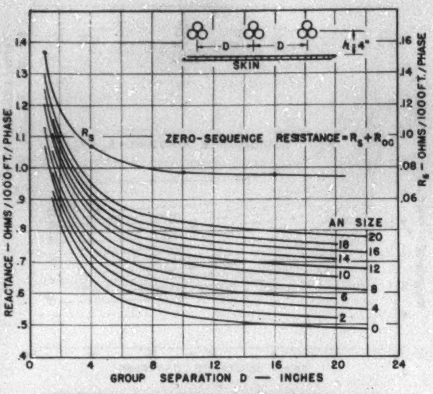
R_{DC} — OHMS / 1000 FT. / PHASE										
AN20	AN18	AN16	AN14	AN12	AN10	AN 8	AN 6	AN 4	AN 2	AN 0
3.42	2.15	1.59	.997	.627	.366	.234	.147	.094	.063	—

Figure 14. Zero-sequence impedance for three flat groups 1 inch above skin at 400 cycles per second



R_{DC} — OHMS / 1000 FT. / PHASE										
AN20	AN18	AN16	AN14	AN12	AN10	AN 8	AN 6	AN 4	AN 2	AN 0
3.42	2.15	1.59	.997	.627	.366	.234	.147	.094	.063	—

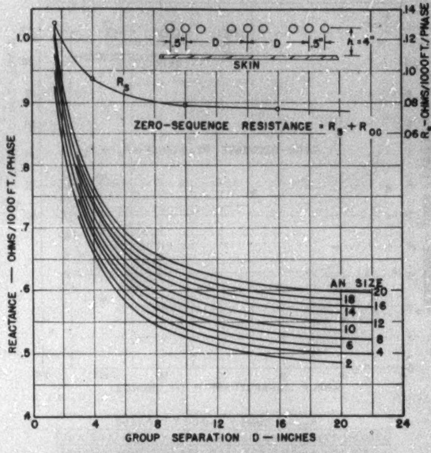
Figure 16. Zero-sequence impedance for three flat groups 3 inches above skin at 400 cycles per second



R_{OC} – OHMS / 1000 FT. / PHASE										
AN20	AN18	AN16	AN14	AN12	AN10	AN 8	AN 6	AN 4	AN 2	AN 0
3.42	2.15	1.59	.997	.627	.366	.234	.147	.094	.063	.042

Figure 17 (above). Zero-sequence impedance for three laced groups 4 inches above skin at 400 cycles per second

Figure 18 (right). Zero-sequence impedance for three flat groups 4 inches above skin at 400 cycles per second



R_{OC} — OHMS / 1000 FT. / PHASE										
AN20	AN18	AN16	AN14	AN12	AN10	AN 8	AN 6	AN 4	AN 2	AN0
3.42	2.15	1.59	.997	.627	.366	.234	.147	.094	.063	—

is also true for groups in 0.5-inch flat spacing where the group separation exceeds about 5 inches (see Figures 9 and 10).

Equations for configurations of four and five 3-phase groups can be developed by the same procedure. Appendix II summarizes the equations developed for GMD and GMR of the configurations considered in this study.

Appendix II. GMD and GMR Equations

Positive-Sequence Equations

Three flat groups:

$$\text{GMD}=[2^5s^9(D^2-s^2)^5(4D^2-s^2)^2\times(D^2-4s^2)^2]^{1/27}$$

$$\text{GMR}=1.166d_s{}^{1/3}D^{2/3}$$

Four flat groups:

$$\text{GMD}=[2^8s^{12}(D^2-s^2)^8(4D^2-s^2)^4(D^2-4s^2)^2\times(9D^2-s^2)^2(9D^2-4s^2)^2]^{1/48}$$

$$\text{GMR}=(3.465d_sD^3)^{1/4}$$

Five flat groups:

$$\text{GMD}=[2^{12}s^{15}(D^2-s^2)^{11}(D^2-4s^2)^4\times(4D^2-s^2)^7(9D^2-s^2)^4\times(9D^2-4s^2)^2(16D^2-s^2)^2]^{1/75}$$

$$\text{GMR}=(9.64d_sD^4)^{1/5}$$

Zero-Sequence Equations

Three groups (laced or flat):

$$\text{GMD}=[(2h)^3(4h^2+D^2)^2(4h^2+4D^2)]^{1/9}$$

(Laced): $\text{GMR}=1.166d^*{}_s{}^{1/3}D^{2/3}$

$$(\text{Flat})\leftarrow\left[\begin{array}{ll}\text{GMR}=1.32d_s{}^{1/9} & (D=1.5'')\\ \text{GMR}=1.63d_s{}^{1/9} & (D=2'')\\ \text{GMR}=1.166d^*{}_s{}^{1/3}D^{2/3} & (D\geq 3'')\end{array}\right.$$

Four groups (laced or flat):

$$\text{GMD}=[(2h)^4(4h^2+D^2)^3(4h^2+4D^2)^2\times(4h^2+9D^2)]^{1/16}$$

(Laced): $\text{GMR}=(3.465d^*{}_sD^3)^{1/4}$

$$(\text{Flat})\leftarrow\left[\begin{array}{ll}\text{GMR}=1.66d_s{}^{1/12} & (D=1.5'')\\ \text{GMR}=2.10d_s{}^{1/12} & (D=2'')\\ \text{GMR}=(3.465d^*{}_sD^3)^{1/4} & (D\geq 3'')\end{array}\right.$$

Five groups (laced or flat):

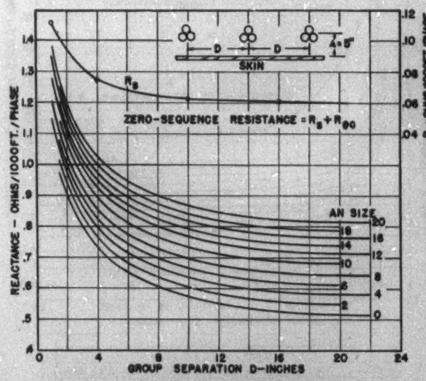
$$\text{GMD}=[(2h)^5(4h^2+D^2)^4(4h^2+4D^2)^3\times(4h^2+9D^2)^2(4h^2+16D^2)]^{1/25}$$

(Laced): $\text{GMR}=(9.64d^*{}_sD^4)^{1/5}$

$$(\text{Flat})\leftarrow\left[\begin{array}{ll}\text{GMR}=1.97d_s{}^{1/15} & (D=1.5'')\\ \text{GMR}=2.53d_s{}^{1/15} & (D=2'')\\ \text{GMR}=(9.64d^*{}_sD^4)^{1/5} & (D\geq 3'')\end{array}\right.$$

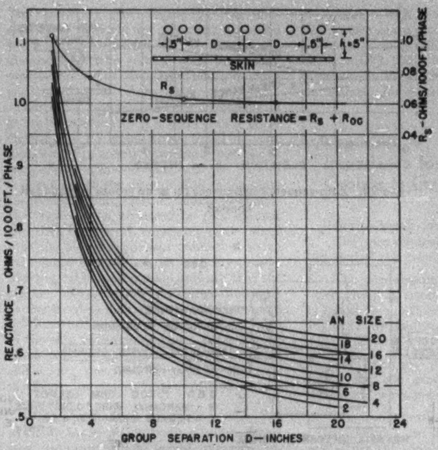
Appendix III. Calculation of Impedance Data

To illustrate the general procedure followed in compiling the data presented in Table II and Figures 9 to 25, several sample calculations will be explained. It is felt that these examples, along with the preceding



R_{OC} — OHMS / 1000 FT. / PHASE										
AN20	AN18	AN16	AN14	AN12	AN10	AN 8	AN 6	AN 4	AN 2	AN0
3.42	2.15	1.59	.997	.627	.366	.234	.147	.094	.063	.042

Figure 19. Zero-sequence impedance for three laced groups 5 inches above skin at 400 cycles per second



R_{OC} — OHMS / 1000 FT. / PHASE										
AN20	AN18	AN16	AN14	AN12	AN10	AN8	AN6	AN4	AN2	AN0
3.42	2.15	1.59	.997	.627	.366	.234	.147	.094	.063	—

Figure 20. Zero-sequence impedance for three flat groups 5 inches above skin at 400 cycles per second

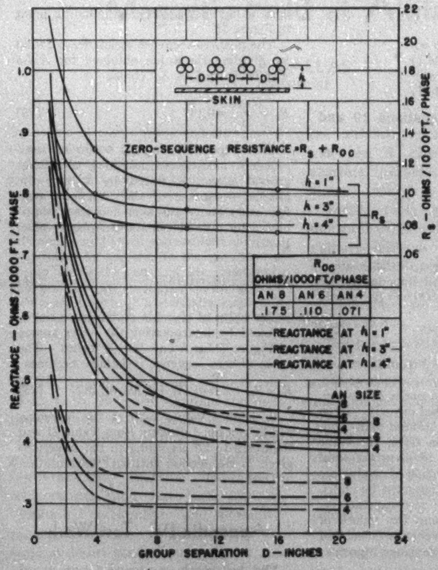


Figure 21. Zero-sequence impedance for four laced groups at 400 cycles per second

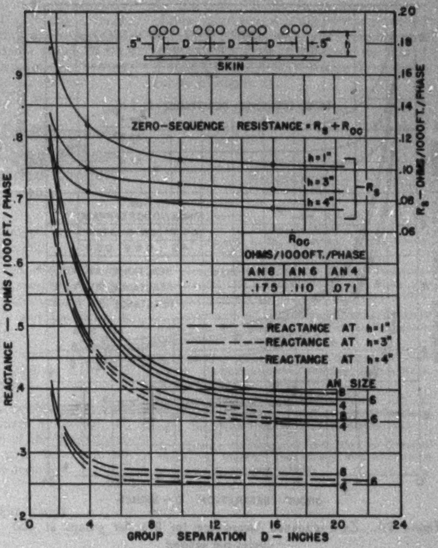


Figure 22. Zero-sequence impedance for four flat groups at 400 cycles per second

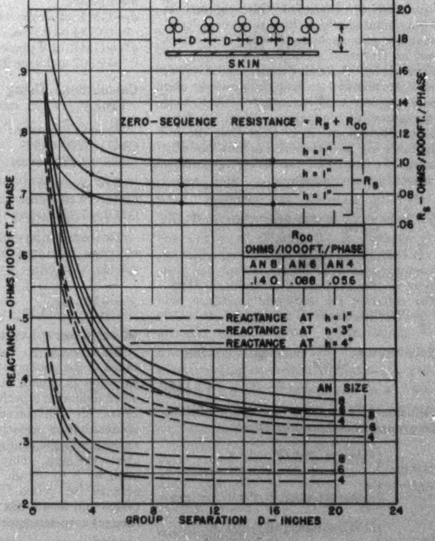


Figure 23. Zero-sequence impedance for five laced groups at 400 cycles per second

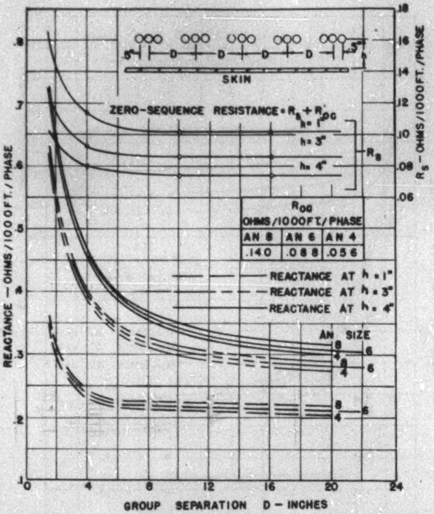


Figure 24. Zero-sequence impedance for five flat groups at 400 cycles per second

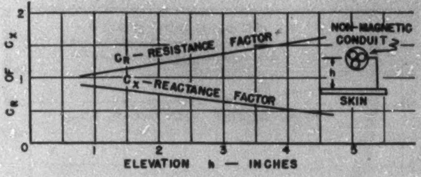


Figure 25. Zero-sequence impedance for laced groups in grounded conduit

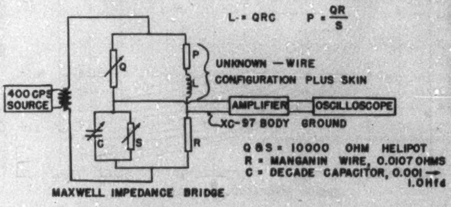

L = QRC
P = QR/S
Q
P
L
UNKNOWN — WIRE CONFIGURATION PLUS SKIN
400 CPS SOURCE
AMPLIFIER
OSCILLOSCOPE
XC-97 BODY GROUND
C
S
R
Q & S = 10000 OHM HELIPOT
R = MANGANIN WIRE, 0.0107 OHMS
C = DECADE CAPACITOR, 0.001 → 1.0 Mfd
MAXWELL IMPEDANCE BRIDGE

Figure 26. Bridge circuit used for impedance measurements

general material, will enable the user of the data to compile additional impedance information as the need arises.

Consider a 3-phase 400-cycle wire configuration consisting of three groups in 0.5-inch flat spacing (three wires per phase) having a group spacing D equal to 4 inches and an elevation h equal to 4 inches; see Figure 1(B). Assuming a wire size of *AN-6*, let it be required to find the positive- and zero-sequence impedance per 1,000 feet of this system of conductors. From the foregoing discussions, it is clear that the inductance can be determined in both cases by using the equation involving geometric mean distances, namely

$$L=\frac{\mu}{2\pi}\ln\frac{\text{GMD}}{\text{GMR}}\text{ henrys per meter} \tag{17}$$

It is desirable to convert this equation to reactance at 400 cycles per second and also to use common instead of natural logarithms. The equation for 400-cycle reactance per 1,000 feet becomes

$$X=0.353\log_{10}\frac{\text{GMD}}{\text{GMR}}\text{ ohms per 1,000 feet} \tag{18}$$

The equations for GMD and GMR of three flat groups from Appendix II are

$$\text{Positive sequence}\begin{cases}\text{GMD}=[2^5s^9(D^2-s^2)^5\times\\ \qquad(4D^2-s^2)^2(D^2-4s^2)^2]^{1/27}\\ \text{GMR}=1.166d_s^{1/3}D^{2/3}\end{cases} \tag{19}$$

$$\text{Zero sequence}\begin{cases}\text{GMD}=[(2h)^3(4h^2+D^2)^2\times\\ \qquad(4h^2+4D^2)]^{1/9}\\ \text{GMR}=[1.166d^*_s{}^{1/3}D^{2/3}]\end{cases} \tag{19A}$$

Data

$s=0.5$ inch
$D=4$ inches
$h=4$ inches
$d_s=$ GMR of single *AN-6* wire
$=0.0734$ inch (see Table I)
$d^*_s=$ GMR of 3-phase group
$=0.307$ inch (see Table I)

Calculations Using Equations 19 and (19A)

Positive sequence:
GMD = 2.50 inches
GMR = 1.232 inches

Zero sequence:
GMD = 9.10 inches
GMR = 1.972 inches

Therefore, from equation 18 the positive-sequence reactance is (all impedances are calculated in ohms per 1,000 feet per phase)

$$X_1=0.353\log_{10}\frac{2.50}{1.23}=0.1085 \tag{20}$$

which checks the value in Figure 9.

With regard to the zero-sequence reactance of the configuration under discussion, equation 18 cannot be used directly. Even when assuming a perfect ground plane and, hence, no skin correction term, the value obtained from equation 18 gives the total reactance of the configuration of conductors above the skin. This value must be multiplied by 3 to obtain the zero-sequence reactance per phase. Using the values of GMD and GMR just given, the calculated (image) zero-sequence impedance becomes

$$X_{0c}=1.059\log_{10}\frac{9.10}{1.97}=0.704 \tag{21}$$

The value of the skin correction term X_s can be found by interpolating the lower curves of Figure 6 for h equal to 4 inches. The value of zero-sequence reactance then becomes

$$X_0=X_{0c}+X_s=0.704-0.019=0.685 \tag{22}$$

This value checks the corresponding point in Figure 18.

The positive-sequence resistance is found by dividing the value obtained for *AN-6* wire in Table I by 3. Hence

$$R_1=\frac{0.440}{3}=0.147 \tag{23}$$

The calculated (image) value of zero-sequence resistance, assuming a perfect ground plane, is found by finding the parallel resistance of the conductor configuration above the skin and then multiplying by 3 to obtain the per-phase value. Hence

$$R_{0c}=\frac{0.440}{9}(3)=0.147 \tag{24}$$

The value of the skin correction term R_s is found by interpolating the lower curves of Figure 5 for h equal to 4 inches. The value of zero-sequence resistance becomes

$$R_0=R_{0c}+R_s=0.147+0.095=0.242 \tag{25}$$

Resistance values such as the foregoing can be obtained directly from Table II and Figures 9 to 25 according to the directions given in the section entitled "Presentation of Data."

Appendix IV. Test Work

The test work was carried out in an *XC-97* body having a skin of *24ST* alloy with an average thickness of 0.04 inch.

This structure (see Figure 7) was considered to be ideal for the test, since the size and distribution of the stiffening members and the general body design are representative of the large aircraft being engineered at the present time.

Three flat groups are shown arranged at a group separation of 1.5 inches and at an elevation of 4.7 inches above the skin. The nine wires are solidly connected together at each end of a 40-foot run, and the far end is grounded to the aircraft skin to simulate the zero-sequence circuit. Each configuration tested was arranged in this manner and then the circuit, consisting of the wires plus the skin return, was connected to an impedance bridge through calibrated leads. Theoretical considerations indicate that the size of wires, within the range normally encountered in aircraft installations, should not have an appreciable effect on the value of the skin correction terms. After several check tests to substantiate this hypothesis, *AN-6* wire was used throughout the tests involving the determination of skin correction terms.

With about 40 feet of clear wire run available in the *XC-97* body, a bridge circuit was required which would measure inductances accurately ranging from 1 to 20 microhenrys and resistance from 0.002 to 0.01 ohm. This represents approximately the values encountered in testing a single wire, three wires, and three 3-phase groups (nine wires). The Maxwell bridge circuit, described in Figure 26, was found to be very satisfactory for these measurements.

Appendix V. Zero-Sequence Impedance for Laced 3-Phase Groups in Grounded Conduit

Refer to Figure 25.

1. The zero-sequence impedance is not a function of the group spacing.

2. The zero-sequence impedance for one laced group in grounded solid-type conduit is found by multiplying the value found in Figure 7 by factors C_R and C_X given in Figure 25.

3. For multiple phase wires the value found for one group is divided by the number of wires per phase.

4. For configurations employing flexible-type conduit, use Figures 4 through 17, neglecting the presence of the flexible conduit. This procedure gives results to approximately 10-per-cent accuracy.

5. *Example:* The zero-sequence-impedance of three laced groups of any wire in grounded solid-type conduit at a group spacing of 4 inches and at an elevation above the skin of 4 inches is found from Figure 7, and the curves in Figure 25:

$$Z_0=(0.557)(1.56)+j(1.68)(0.55)$$
$$=0.90+j0.925 \text{ ohms}$$

References

1. Impedance of 400-Cycle Three-Phase Power Circuits on Large Aircraft and Its Application to Fault-Current Calculations, C. K. Chappuis, L. M. Olmstead. *AIEE Transactions*, volume 63, 1944, pages 1213–19.

2. Principles of Electric Power Transmission (book), L. F. Woodruff. John Wiley and Sons, Inc., New York, N. Y., second edition, 1938.

3. Introduction to Electric Power Systems (book), J. G. Tarboux. International Textbook Company, Scranton, Pa., 1944.

4. General Electric Data Folder Number 63004. General Electric Company, Schenectady, N. Y., January 25, 1943.

5. Symmetrical Components (book), C. F. Wagner, R. D. Evans. McGraw-Hill Book Company, Inc., New York, N. Y., first edition, 1933.

6. Circuit Analysis of A-C Power Systems (book), E. Clarke. John Wiley and Sons, New York, N. Y., volume I, 1943.

No Discussion

Arc Interruption Phenomena in a Magnetic Field at Altitude

J. P. DALLAS
MEMBER AIEE

Synopsis: Arc immobility and arc reversal phenomena limit the use, affect the design, and dictate the test procedure required for aircraft circuit interrupting equipment using magnetic arc suppression. A summary of the literature on this subject is offered. An explanation of the cause of transient arc immobility and small gap arc immobility is proposed. An addition to the consensus of theory as to the cause of arc immobility and arc reversal is suggested usefully relating arc length as a critically important factor to the understanding of these phenomena.

THE aircraft electrical engineer designs to the most rigorous specifications known to the electrical industry and for environmental conditions that would have been regarded as fantastic a few years ago. Extreme conditions of temperature, sand and dust, altitude, vibration, acceleration and humidity, ranging from the environment of the tropical desert sand storm to the stratosphere, are his common problems. While meeting astonishing requirements for the absolute minimum of weight and space he is expected to achieve the absolute maximum of reliability. The aircraft electrical engineer must accomplish these semimiracles because human lives, as well as the completion of military missions on which the safety of the nation may depend, are in turn dependent on the correct functioning of something as "unimportant" as an aircraft switch.

Further to confound the aircraft switch designer, the very laws of physics which earth-bound electrical designers deal with unquestioningly are found to require re-examination. It had been assumed for more than 50 years that an electric arc, like any other conductor free to move in a magnetic field, would move in a direction determined by Ampere's law. This assumption is valid for all practical purposes of circuit breaking equipment operating at sea level, but something happens when Ampere's law is applied to similar equipment at altitude. As the altitude is increased, an electric arc in a magnetic field may falter, stop, and finally reverse its direction of motion.[2,3,9,12,13]

Probably nothing can be more disconcerting to the designer or to those responsible for the testing of equipment than an abrupt discontinuity in characteristics. Pity the poor designer who finds that a switch which will interrupt a large current successfully will not interrupt 10 per cent of that value. That is exactly the type of discontinuous behavior and failure which may be expected with magnetic arc suppression when operated at reduced pressure. This designer's dilemma is illustrated in the curve of Figure 3. A low-travel snap-action switch with magnetic arc suppression is shown to open consistently and stably an inductive 120-volt d-c circuit of 9 amperes from sea level to an altitude of 25,000 feet while failing to interrupt 1 ampere. This disconcerting discontinuity in switch performance was caused by a phenomenon known as "arc immobility" which is a precondition to "arc reversal." It is, of course, ap-

---

Paper 52-333, recommended by the AIEE Air Transportation Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Middle Eastern District Meeting, Toledo, Ohio, October 28–30, 1952. Manuscript submitted August 6, 1952; made available for printing September 9, 1952.

J. P. Dallas is at 8511 Vicksburg Avenue, Los Angeles, Calif.

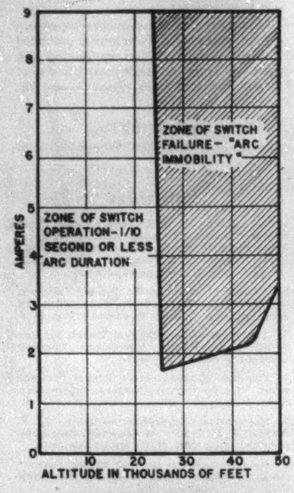


Figure 1. 0.070-inch contact gap, 300-gauss permanent magnet field

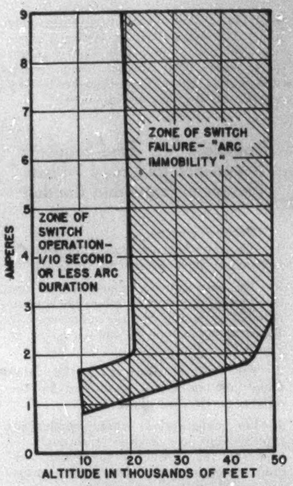


Figure 2. 0.035-inch contact gap, 300-gauss permanent magnet field

parent that arc immobility is a fancy way of saying that a magnetic arc suppressor has failed completely. Reversal is merely the completion of the process in which two opposing forces, acting on the arc, balance each other at immobility and interchange their relative strengths at reversal.

Arc immobility and reversal phenomena, because they are largely tied to reduced pressure operation of switchgear, have been of academic interest to the electrical industry and of primary importance to the aircraft industry where reduced air pressure is a normal ambient condition. To the aircraft electrical designer magnetic arc suppression is in general the lightest and simplest means so far available. While it is true that elementary switch design suffices for low powers and low voltages, it should be remembered that high-voltage conditions must be dealt with on all airplanes today.

Arc suppression problems do not occur only on airplane systems using voltages higher than 28 volts. It is well to recall that among the most essential elements in a standard 28-volt aircraft system are the reverse current and/or main generator contactors. In the present 28-volt systems these contactors can be called on to interrupt more than 100 volts under conditions of overvoltage. Due to the design of present generator terminal arrangements and the probable hazards at fire wall and nacelle terminals, short circuits of the field to the positive line giving full field output voltages in excess of 100 volts are matters of record. The fact is that current designs of 28-volt generator contactor equipment are depending on magnetic arc suppression to meet these conditions.

For these reasons consideration of the failure and design liabilities of this type of arc interrupting device from the aircraft designer's point of view is desirable.

Arc Immobility

The following experimental data resulted from operating at various altitudes a small snap action aircraft limit switch with magnetic arc suppressor.

Procedure

Referring to Figures 1, 2, and 3, a limit switch, normally rated at 10 amperes for 120 volts d-c inductive circuits, was modified to allow for adjusting the gap to the various values indicated. This switch made use of an Alnico permanent magnet located to produce a transverse field of 300 to 500 gausses in the contact gap. The switch was arranged for mechanical operation in an altitude chamber with provisions for recording the results photographically.

The ampere values shown on these curves are the maximum currents which could be consistently interrupted without exceeding an arc duration of 1/10 second. These values are far in excess of practical ratings. One-tenth second arc duration operation is in this case very close, probably within 15 per cent, of the current at which the switch would completely fail to open the circuit.

For each switch gap value 5,000-foot increment changes of altitude were made from sea level to a 50,000-foot simulated altitude. At each change in altitude the current was raised until the arc time exceeded 1/10 second in duration and then dropped back in 1 percent steps until repeated operation for 20 breaks could be secured with a consistent arc time of approximately 1/10 second.

An inductive load with an inductance-reactance ratio of approximately 0.02 second was used. This time constant was an actual value determined from oscilloscope observation of the current build-up time in the test circuit.

Results

Referring to Figure 1 in which the test sample was adjusted for a contact gap of 0.070 inch, attention is directed to the following:

1. The 0.070-inch gap switch interrupted 9 amperes of 120 volts d-c inductive load from sea level to approximately 25,000 feet.
2. At 25,000 feet the arc became diffuse, blowout action of the magnetic field on the arc became uncertain, the arc was immobilized, and the interrupting capacity of the switch precipitiously dropped to $1^{3}/_{4}$ amperes.
3. Above the 25,000 foot altitude drop there was a gradual increase in the capacity of the switch to a value of $3^{1}/_{2}$ amperes at 50,000 feet.

Referring to Figure 2 in which the test sample was adjusted for a contact gap of 0.035 inch, note the following:

1. A discontinuity appears in this figure. It is to be noted that between 10,000 and 20,000 feet the switch will not interrupt values around 1 ampere while interrupting values from 2 to 9 amperes. This discontinuity in interrupting capacity caused by arc immobility is the most significant fact demonstrated in this series of tests.
2. The 0.035-inch contact gap switch interrupted 9 amperes from sea level to 20,000 feet of 120-volt inductive load.
3. At 20,000 feet the arc became diffuse, blowout action of the magnetic field on the arc became uncertain, the arc was immobilized, and the interrupting capacity of the switch precipitously dropped to $1^{1}/_{4}$ amperes.
4. Beyond the 10,000-foot discontinuity, the interrupting capacity of the 0.035-inch contact gap switch rose to a value of $3^{1}/_{2}$ amperes at 50,000 feet.

Referring to Figure 3 in which the test sample was adjusted for a contact gap of 0.020 inch, attention is directed to the following:

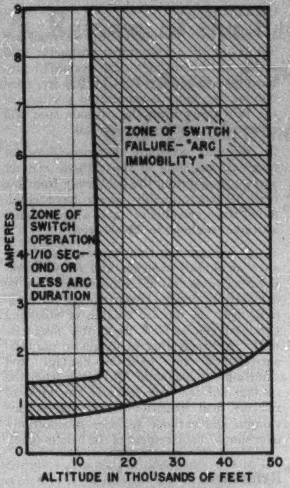


Figure 3. 0.020-inch contact gap, 300-gauss permanent magnet field

1. The discontinuity in interrupting capacity of the switch which appeared at 10,000 feet in Figure 2 has been extended to sea level. Here again is the condition previously mentioned where the switch at sea level to 15,000 feet will not interrupt values around 1 ampere but will interrupt 9 amperes.

2. At 15,000 feet the arc became diffuse, blowout action of the magnetic field on the arc became uncertain, and the arc was immobilized.

3. The interrupting capacity of the switch rises from the discontinuity just mentioned to a value of $2^{1}/_{4}$ amperes at 50,000 feet.

Comment and Conclusions

The problem of adequately testing magnetic arc suppression circuit interrupting devices is complicated by the possibility of discontinuities in maximum current interrupting characteristics such as have been noted in Figures 2 and 3. It may be necessary to test such equipment at small intervals of altitude change and current values for the complete range through which the equipment is expected to operate.

In the technical discussions listed in references 1 through 13 there is only limited agreement on the cause of arc immobility. The consensus of the investigators may be summarized thus:

1. Arc immobility and reversal in a transverse magnetic field are the result of two opposing forces, one acting on the arc plasma or column, and the other in the opposite direction on the cathode spot.[3]

2. The transverse magnetic field exerts a force on the arc column of $F=IB$, where I equals the arc current and B is the magnetic field. This force on the arc plasma or column does not change with pressure and is exerted in a normal direction which would cause the arc to move in the same manner as any other conductor would be expected to move and in accordance with Ampere's law.[9]

3. The transverse magnetic field under certain conditions of pressure, cathode material, and arc atmosphere constituency exerts a force on the cathode spot in an opposite direction to that on the arc column and in opposition to the direction predicted by Ampere's law.[9,13]

4. The reverse driving force on the cathode spot is increased by reduction of pressure (altitude increase) and by an increase in the transverse magnetic field.[3]

The principle disagreement of the referenced investigators is concerning what happens in the cathode spot region to cause reverse forces and/or motion. Speculation is substituted for fact, and all we seem to have in the final analysis are watered-down mathematical assumptions. The problem is difficult. Since the cathode drop region usually is defined as being of the order of the mean free path of an electron emitted from the cathode, and considering that cathode spot current densities are estimated at 100,000 amperes per centimeters squared, and the temperature at 4,000 degrees Kelvin, it is easy to understand why so little is known.

Starting with the tangibles of observed arc column movement and cathode spot reverse movement, outlined in the foregoing as the consensus of the references, a simple explanation of the discontinuity in the maximum current interrupting characteristics of Figures 2 and 3 may be advanced. If the force acting on the arc column is $F=IBL$ it is apparent that decreasing I, the arc current, or L, the arc length, would reduce this force tending to drive the arc in a normal direction. If such reduction of the normal arc column driving force results in immobilizing the arc at some current value above the interrupting capacity of the simple switchgap, it is obvious that a discontinuity in the maximum current interrupting capacity characteristic would occur. For example, referring to Figure 3, the lower limit of the 1-ampere sea level to 15,000-foot discontinuity may be regarded as occurring at approximately 3/4 ampere and representing the arc interrupting capacity of the switchgap without magnetic arc suppression. The upper limit of the discontinuity represented by a line at 1.4 amperes from sea level to 15,000 feet is roughly the point at which the magnetic arc suppression starts to be effective because the increased I has increased the force acting on the arc column sufficiently to produce arc motion in a normal direction.

Yamamura, in discussing immobility phenomena and reverse driving phenomena of the electric arc,[13] considers at some length the arc immobility effect of small gaps. He also considers the transient immobility or delay time between the start of switch opening and the first motion of the arc. He states that these effects may be due to thermal conditions and implies that higher temperatures will be present and are the cause. Since both small gap immobility and switch opening transient immobility are commonly observed in aircraft circuit interrupting devices under low-current and arc temperature conditions, this explanation is incomplete if not doubtful. It is suggested, therefore, that both of these phenomena are the result of low (normal direction) driving forces on the arc column due to its short length. It is obvious if the driving forces on the arc column are $F=IBL$ that L, the arc length, must reach some substantial value before the driving forces on the arc column can produce motion. The time interval, then, between the start of switch opening and the time the arc length is sufficient to produce effective arc column driving force, is the transient immobility time. Small gap immobility is the case where the total arc length never reaches a value sufficient to cause

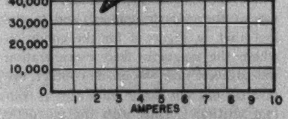


Figure 4. Magnet field concentrated near the anode, 0.10-inch contact gap, 0.02 time constant of inductive load

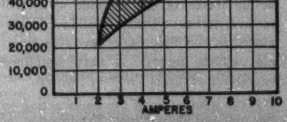


Figure 5. Magnet field concentrated near the cathode, 0.10-inch contact gap, 0.02 time constant of inductive load

effective driving forces to be applied to the arc column, which, of course, may be restrained by small amounts of reverse driving force on the cathode spot.

Of course, discontinuity in the maximum current interrupting characteristics of electromagnetic arc suppression devices is increased by the simultaneous reduction in both field and arc current values. Discontinuities in the characteristics of this type of equipment, however, are relatively better known. It has been the purpose here to call attention to the existence of such critical problems in permanent-magnet or constant field arc interrupting devices now common in the aircraft industry.

It would be wrong to assume, without a better understanding of arc immobility phenomena at altitude, that discontinuities in the current interrupting characteristic of permanent-magnet arc suppression devices are confined to small gaps of the order of those discussed, 0.050 inch. Increase of permanent-magnetic arc suppressing field strengths from the 300 gausses of the experimental switches considered in this paper to the several thousand gauss field densities now in use in certain aircraft contactors will shift the conditions for current interrupting discontinuities to much larger gaps than those considered here. This may be deduced from the fact that increased field strengths have been shown to increase the tendency toward arc immobility and to make possible arc immobility at larger gaps when other conditions are constant.[13]

Arc Reversal

Michael Faraday[1] in his published diary for January 13, 1862, described the separation of an electric arc into two parts by a transverse magnetic field, which were acted upon differently. The appended references as well as the present discussion could be regarded as adding more detail to Faraday's original observation.

The consensus of the references is that an arc immobilized in a magnetic field by a near balance of the forces acting on the arc column in compliance with Ampere's law, and the oppositely acting forces of the cathode spot, will reverse motion when the balance is displaced in the direction of the reverse acting cathode forces. A very simplified expression of this is $F=P-C$, where

F is the summation of the forces acting on the arc and determining its direction of motion

P is the positive-acting force on the arc plasma or column

C represents the negatively acting forces of the cathode drop region

Thus immobility is the condition where P and C are approximately equivalent, and reversal is the condition where C effectively exceeds P.

PROCEDURE

Quill and Rader[3] described a device for investigating arc reversal in a magnetic field, consisting of two curved copper bars mounted to make contact at their centers and arranged so that the arc could move horizontally equally well in either direction. A separately excited field magnet supplied a transverse field to the gap between the two curved bar contacts. The following experimental data resulted from constructing a similar device modified to provide adjustable contact opening of the upper curved bar contact actuated by a solenoid. The position of the transverse electromagnetic field pole pieces was adjustable vertically so that the strongest portions of the field could be concentrated in the region of either the cathode or anode contact.

The modified Quill-Rader device was installed in an altitude chamber and arranged to be operated with external controls and provisions for recording the result. Gap openings from 0.02 to 0.30 inch were investigated at field densities from 200 to 3,000 gausses. The narrow field pole faces were adjusted for three different positions, 0.075 inch below the cathode contact surface, 0.075 inch below the anode contact surface, and on the center line of the gap.

Displacing the field magnet vertically 1/8th inch altered the relative field densities on the arc column and cathode contact surface only a small amount. While measurements were not made it is estimated that the change in relative field densities was of the order of 20 per cent.

RESULTS

Typical results are shown in Figures 4 and 5 for contact gaps of 0.10 inch operating is a 120-volt d-c inductive circuit with a time constant of 0.02 second. Attention is directed to the start of arc reversal at 23,000-foot altitude in Figure 5 where the magnetic field concentration was at the cathode contact, and to the fact that in Figure 4 for the condition where the magnetic field concentration was greatest near the anode contact, arc reversal did not occur until the 37,000-foot altitude. Thus with the magnetic field concentrated near the anode the altitude was raised 14,000 feet higher before arc reversal occurred.

CONCLUSIONS

1. Since arc immobility and reversal are the result of a transverse magnetic field acting on the cathode drop region to produce reverse-acting forces on the arc, then moving the more concentrated portions of the field away from the cathode spot will reduce the tendency for reversal.

2. We may speculate that provisions for magnetic shielding of the cathode contact surface or shunting the field away from the cathode area probably would have eliminated the reversal effect up to a much higher altitude—perhaps to 50,000 feet.

3. It is suggested that improved performance of aircraft switches and relays incorporating magnetic arc suppression may be achieved by designs concentrating the magnetic field on the arc column or shielding the cathode surface from the magnetic field. A simple ring of magnetic material surrounding the cathode contact and shunting the magnetic field away from the contact surface may increase the ratio of relative field densities acting on the arc column and cathode surfaces substantially over simple displacement of the field magnet.

References

1. MICHAEL FARADAY'S DIARY (book). Volume VII, 1820–1862, January 13, 1862.

2. THE MERCURY ARC CATHODE, **Charles G. Smith.** *Physical Review* (New York, N. Y.), July 1942.

3. D-C ARC INTERRUPTION FOR AIRCRAFT, **J. S. Quill, L. T. Rader.** *AIEE Transactions* (*Electrical Engineering*), volume 63, December 1944, pages 883–88.

4. PEAK VOLTAGES WITH D-C ARC INTERRUPTION FOR AIRCRAFT, **Virgel E. Phillips, Walter P. Mitchel.** *AIEE Transactions* (*Electrical Engineering*), volume 63, December 1944, pages 944–49.

5. A NOTE CONCERNING THE MOTION OF ARC CATHODE SPOTS IN MAGNETIC FIELDS, **R. L. Longini.** *Physical Review* (New York, N. Y.), volume 71, May 1947, page 642.

6. MOTION OF LOW PRESSURE ARC SPOTS IN MAGNETIC FIELDS, **R. L. Longini.** *Physical Review* (New York, N. Y.), volume 72, July 1947, page 184.

7. ARC MOTION REVERSAL IN TRANSVERSE MAGNETIC FIELD BY HEATING CATHODE, **Charles G. Smith.** *Physical Review* (New York, N. Y.), volume 73, March 1948, page 543.

8. CURRENT DENSITY OF THE ARC CATHODE SPOT, **J. D. Cobine, C. J. Gallagher.** *Physical Review* (New York, N. Y.), volume 74, November 1948.

9. THE REVERSE BLOWOUT EFFECT, **Gary J. Himler, George I. Cohn.** *Electrical Engineering*, volume 67, number 12, December 1948, pages 1148–52.

10. REVERSE BLOWOUT EFFECT, **C. J. Gallagher, J. D. Cobine.** *Electrical Engineering*, volume 68, number 5, May 1949, page 469.

11. ARCS BETWEEN MOVING ELECTRODES, **W. B. Kouwenhoven, T. B. Jones.** *Electrical Engineering*, volume 68, number 10, October 1949, page 834.

12. THE RETROGRADE MOTION OF THE ARC CATHODE SPOT, **C. J. Gallagher.** *Journal of Applied Physics* (New York, N. Y.), volume 21, August 1950.

13. IMMOBILITY PHENOMENA AND REVERSE DRIVING PHENOMENA OF THE ELECTRIC ARC, **Sakae Yamamura.** *Journal of Applied Physics* (New York, N. Y.), volume 21, March 1950.

No Discussion

A Review of the Factors Affecting the Temperature of Airborne Electric and Electronic Equipment

LEONARD J. LYONS
NONMEMBER AIEE

Synopsis: Continuously increasing airplane performance has outstripped the ability of electric and electronic equipment to withstand environmental temperature conditions without special treatment. The problem is one of controlling the actual temperature of the equipment to a value that will provide satisfactory life. The equipment temperature is affected by many factors, including such things as ambient air density and orientation of the unit with respect to surrounding surfaces. The common specification of ambient air temperature is no longer adequate, and it is necessary to work with actual component temperatures.

THE PROBLEM of maintaining a satisfactory environment for airborne equipment is becoming increasingly difficult as the quantity of equipment increases and as airplane and engine performance is extended. The output of the various heat sources, such as hot engine surfaces, heat dissipating electric and electronic equipment, and aerodynamic heating has increased significantly every year. At the same time, increased airspeeds and altitudes reduce the usefulness of ambient air as a heat sink.

This discussion has two primary objectives: First, to review some of the basic factors involved in the transfer of heat from electronic and electric apparatus to its environment, and second, to illustrate the significant and apparently often unexpected effects of the environment on the operating temperatures of the apparatus.

Considerable complication in the design of equipment conditioning systems results from the fact that ambient air temperature long has been utilized as a primary reference in the qualification of equipment at high temperatures. Most qualification programs and specifications call for the unit to be subjected to a certain ambient temperature, and it has been considered to be adequate "conditioning" if the air temperature around the unit, as installed in the airplane, is held below this value. In actual fact, the only temperature that is significant in determining whether or not a given unit will operate satisfactorily is the temperature of the unit itself, and in many installations the ambient air temperature has only a secondary effect on the unit temperature.

Obviously, the actual component temperature is the value that is of primary concern. Solder starts to melt at a certain temperature level, independent of the cause or origin of the temperature; it can be internal power dissipation or it can be a result of heat received by radiation alone, but the solder still will melt at a given temperature. The same thing applies to any other type of physical deterioration, such as the charring of insulation, or the softening of impregnants. Similarly, electrical characteristics are a function of the temperature of the material; it is the actual wire, plate, or mica temperature that counts, not ambient air temperature. Of course, actual material failure temperatures do not even have to be reached; as they are approached, the life of the components drop off rapidly and it is necessary to stay well below the temperatures at which melting and burning actually start.

Aside from internal heat dissipation, the most important factor in establishing component temperature is frequently the temperature of the surrounding surfaces, such as the airplane structure or other high-temperature components. In a test setup, the temperature of the test cell walls may be most significant in determining the equipment temperature. A situation could readily occur where a piece of inert equipment is qualified in a test box at an ambient air temperature of 160 degrees Fahrenheit, while the actual unit temperature is only 150 degrees Fahrenheit. Furthermore, the same equipment may be installed in the airplane where the ambient temperature is still 160 degrees Fahrenheit, and actually be operating at a temperature of 170 degrees Fahrenheit. Temperature discrepancies such as this easily could be the difference between satisfactory operation during the test and completely unsatisfactory operation in the airplane. A more detailed illustration of these effects will be given a little later in the paper.

At this point, a brief review of heat transfer calculations will help to illustrate how the equipment temperatures are affected by the environment.

Any elementary course or text on the subject of heat transfer will provide enough information to set up the following equation for a heat balance on an assumed component.

$$Q=(T_s-T_a)h_cA_c+(T_s-T_w)h_rA_r$$

This equation describes the flow of heat from a component that is dissipating power within itself, and for any given operating conditions, will permit calculation of the temperature of the component surface. In this equation Q is the power dissipated (watts in electrical language or Btu per hour in thermal units).

The first group of terms on the right expresses the heat that is lost to the air by convection, T_s and T_a being the equipment surface and the ambient air temperatures respectively; h_c being the convection coefficient, giving the heat transfer in British thermal units per hour per unit area for 1-degree temperature difference; and A_c being the surface area over which the temperature difference and the convection coefficient are effective. It must be noted that this simplified expression has a limited application. In particular, modification would be required before it could be used for relatively high air velocities. Of particular significance is the fact that h_c is a function of many other factors; it is affected primarily by air velocity, and to lesser degrees by such things as air density, temperature, and the size and surface condition of the component. Also it is affected by the orientation of the component, as is the effective value of A_c. Difficulties also can be encountered in the choice of proper values of T_a which, for a part inside a chassis inside an airplane, is not outside air temperature, nor even compartment air temperature. A realization of this fact—that the ambient air temperature is not always what it seems to be—has confused many electrical and electronic installation men and designers. In general, the ambient air temperature in a chosen area must be calculated by another separate heat balance on the mass of the air, except for those rare cases when

---

Paper 52-335, recommended by the AIEE Air Transportation Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Middle Eastern District Meeting, Toledo, Ohio, October 28–30, 1952. Manuscript submitted August 6, 1952; made available for printing September 5, 1952.

LEONARD J. LYONS is with Northrop Aircraft, Inc., Hawthorne, Calif.

the rate of airflow through the area is so great that the air suffers no significant rise over inlet temperature. A simultaneous solution of the two heat balance equations for the equipment surface and ambient air temperatures must be made.

Similarly, the second group of terms on the right side expresses the heat that is lost by radiation to the walls surrounding the component. Frequently, the temperature of these surrounding walls is established primarily by factors outside of the equipment compartment, such as external skin boundary layer temperatures, or heat from an engine section. When the wall temperature is affected significantly by radiation from the enclosed equipment or by convection from the compartment air, then another heat balance on the walls themselves must be spelled out and the three equations solved simultaneously for equipment, air, and wall temperature.

This expression also has been considerably simplified. Heat transfer by radiation is actually a function of the difference between the fourth powers of the absolute surface temperatures. In the form shown, this is accounted for by the value of h_r, which is therefore a function of both T_e and T_w. Furthermore, h_r includes an allowance for the emissivity of the surfaces involved, which may be a factor anywhere between 0.95 for painted surfaces and 0.05 for highly polished surfaces. Also included is another more complex factor which allows for the relative shapes and orientation of the surfaces. It should be noted also that the effective radiation area often can be quite different than the convection area. For example, on a finned cylinder, A_r is much smaller than A_c, and the determination of proper values for these areas is often a difficult problem.

For simplification, the complete equation ignores heat transfer by conduction, which could be represented by a third term on the right. In most current installations conduction is a small factor, although with careful design it can be utilized to bring about significant reductions in equipment temperatures.

Rearrangement of the equation to solve for equipment temperatures gives

$$T_e = \frac{Q + T_a h_c A_c + T_w h_r A_r}{h_c A_c + h_r A_r}$$

It can be seen now that the equipment temperature is a function of a great many factors, including these significant ones:

1. Power dissipated by the unit.
2. Ambient air temperature.
3. Ambient air velocity.
4. Ambient air density.
5. Size, shape, orientation, and surface condition of the unit.
6. Size, shape, orientation, and surface condition of the surrounding surfaces.
7. Temperatures of surrounding surfaces.

Fortunately, the values of A_e and A_r usually are not affected greatly by the selection of a given set of operating conditions. However, all the other factors that affect equipment temperatures are definitely a function of the selected operating conditions, and ambient air temperature is only one of these factors.

It should be noted that this equation for equipment temperature is equally applicable when the heat dissipation Q is equal to zero. In this case, T_e is obviously at some value between T_a and T_w, and the problem is one of protecting the equipment from its environment. The heat transfer coefficients from the hot environmental elements must be reduced, and the coefficients to the cooler elements increased. The problem is different when Q is relatively large, and the highest possible heat transfer rate is desired. Thus in one installation it may be best to paint a unit black, while in another the same unit should be given a shiny finish.

Table I illustrates the effect of some environmental conditions on an assumed unit, 1 inch in diameter and 1 inch long, dissipating only 1 watt, under various conditions.

Case 1 represents bench test operation with the room air and walls at 60 degrees Fahrenheit and the unit freely supported so as to permit natural convection to take place. The moderate power loss of 1 watt results in only a 43-degree-Fahrenheit temperature rise above ambient, and almost one-half of the heat transfer is by radiation.

Case 2 represents conditions in an ideal test box, conforming to the apparent intent of the Air Force specifications, where the air and wall temperatures are maintained at 160 degrees Fahrenheit. For the 100-degree-Fahrenheit increase in environmental temperature, the unit temperature has only increased 92 degrees Fahrenheit. This small gain of 8 degrees Fahrenheit is a result of the increased radiation coefficients at the higher absolute values of temperature.

Case 3 represents conditions in an assumed typical test box where, because of relatively low temperatures—such as 60 degrees Fahrenheit—in the room outside of the box, the inside walls of the box despite insulation are only 130 degrees Fahrenheit, while the air inside is at the specification value of 160 degrees Fahrenheit. It can be seen that the unit temperature is 13 degrees Fahrenheit less than in an ideal box.

In case 4, another typical test condition is assumed, where the box air temperature is maintained by gentle circulation of the air, presumably through a heat exchanger outside of the box. An air velocity of only 6.3 feet per second was assumed, yet this is sufficient to drop the equipment temperature another 11 degrees Fahrenheit, making it a total of 24 degrees Fahrenheit below the ideal qualification condition.

Now, in case 5, a typical airplane installation is analyzed. Commonly, forced draft is not provided and natural convection coefficients apply. A surrounding surface temperature of 180 degrees Fahrenheit is assumed, which is not at all pessimistic for a unit surrounded by other heat-producing devices. It can be seen that the unit temperature has increased to 205 degrees Fahrenheit, and is 25 or 30 degrees Fahrenheit hotter than in typical qualification conditions.

In case 6, this installation has been taken to an altitude of 50,000 feet, where the low air pressure reduces the convection coefficient sufficiently to raise the unit temperature an additional 24 degrees Fahrenheit and 85 per cent of the heat transfer is by radiation.

Case 7, illustrating the effect of the cur-

Table I. Computed Surface Temperature T_e

			Typical Test		Typical Installation		
	Bench, Case 1	Ideal Test, Case 2	Natural Convection, Case 3	Forced Convection, Case 4	Sea Level, Case 5	50,000 Feet, Case 6	Sea Level, 2 Watts, Case 7
T_e, degrees Fahrenheit	103	195	182	171	205	229	235
Per cent radiation	44.5	54.4	76.3	54.8	41	84.5	42.2
Per cent convection	55.5	45.6	23.7	45.2	59	15.5	57.8
T_a, degrees Fahrenheit	60	160	160	160	160	160	160
T_w, degrees Fahrenheit	60	160	130	130	180	180	180

rent miniaturization program, is for a typical installation at sea level, with a power dissipation of 2 watts instead of 1. It can be seen 1 more watt of power has raised the unit temperature by an additional 30 degrees Fahrenheit, and this is typical of the problems being encountered daily in the cooling of electric equipment.

This also points out the necessity for having reliable data on the actual power dissipation of the equipment. Sometimes it is extremely difficult to secure this information from the designers of the equipment, and this results in considerable confusion in the design of the necessary cooling systems. It is frequently discovered that the numbers obtained are inputs, sometimes outputs, and sometimes maximum catalogue ratings bearing no relationship to the use in a given installation. It cannot be emphasized too strongly that, although it is sometimes very difficult to estimate the electrical efficiency of a design, procedures must be devised to measure the input and output so that the losses are known. It is understood that in many electronic problems actual efficiency is often, of necessity, a secondary problem. Requests for this information are not made with any wish to embarrass, but it is essential that the numbers be available no matter what they are. However, it was startling to learn on a certain airplane that the total of the heat that had to be disposed of from the various equipment compartments was about 75 per cent of the power available from the generators. Although the dissipations may be off by as much as 25 or 50 per cent, the indications are that equipment temperatures can be reduced very effectively if ways can be found to reduce this wasted power.

It has been illustrated that a specification of ambient temperature is no longer an adequate means of insuring satisfactory high temperature operation, and that actual equipment temperatures must be specified and controlled. An unfortunate result of past reliance upon ambient temperature measurements is the fact that only very limited amounts of data are available concerning what temperatures the various pieces of equipment are actually capable of withstanding. Such data are obviously necessary before an intelligent conditioning system can be designed. Furthermore, these data must be expressed in terms of life versus temperature, so that each installation and type of application can be evaluated properly. It has been extremely difficult to get this fact across to those concerned with the specification, development, and testing of equipment. Ambient temperature data are only of value in assisting to evaluate the heat transfer mechanisms and factors for a given installation, and they give an inaccurate picture if the other environmental conditions are not taken to account.

The example has been restricted to that of a single, rather elementary unit in relatively uniform surroundings, and unfortunately most practical problems are not nearly this simple. Common mountings, such as on a flat surface, usually do not permit taking advantage of cooling by conduction to the surface, but turn the estimation of coefficients and areas for convection and radiation into a major problem. Obviously, the surface temperature is distinctly nonuniform, and it must be assumed that the hottest point on the unit is the one that determines its useful life.

Most of the more complex assemblies are almost impossible to analyze intelligently. They are frequently supplied to the airframe manufacturer as government-furnished equipment, in which case he has no control over, and little knowledge of, the internal arrangements. Practically nothing useful can be done to insure satisfactory temperatures, since these units are controlled by specifications that define ambient air temperature and nothing else. Some attempts may be made to calculate the temperatures of the box surfaces that enclose the assemblies and then distribute the ventilating air where it seems most needed, but there are absolutely no valid criteria for determining the amount of air needed, or how it can be most efficiently utilized.

The temperatures of these boxes are nonuniform, even in an ideal environment, due to nonuniform internal power distributions. Furthermore, each installation of a given box will result in different distributions of the surface temperature. An assembly designed with all the heat dissipating units in one end normally will be hot at that end, and the other end of the box will be relatively cool. But, install that box with the cool end facing another hot surface, and those inert units that were previously safe immediately become critical.

Current equipment installations are in a most difficult position, since the equipment apparently can not stand the high environmental temperatures that are to be expected, unless flight durations are short enough to permit taking advantage of the thermal lag of the components. Data to substantiate this line of attack are very limited, and scientific application of such data is actually a very complex and time-consuming procedure. However, the alternatives are equally undesirable; either restrict the airplane performance, or install actual refrigeration systems.

For elementary components, presentation of a curve of maximum allowable surface temperature versus life would put the onus on the installation man to provide an environment giving a surface temperature low enough to give the necessary life. If expensive procedures are necessary, such as airplane flight restriction or the use of refrigeration, then design criteria are available that will admit no argument. However, for the more complex equipment, it would be necessary to establish curves for each of the potentially critical surfaces. Considerable effort and more experience along this line are necessary before a really satisfactory procedure can be set up.

Within a few years it can be hoped that improved materials will become generally available and that they will serve to permit an increase in actual operating temperatures and, presumably, in the specification temperatures. However, benefits from such developments probably will not be sufficient to provide us with efficient installations on future high-speed aircraft. The big improvements will come about when the designers of the equipment become constantly aware of the magnitude of the problem and either acquire the necessary knowledge themselves or secure the assistance of heat transfer specialists. Units will be designed with improved electrical efficiency, and with careful internal arrangement to permit optimum heat flow to planned areas of heat rejection. Significant but isolated steps in this direction already have been taken, and some remarkable results have been obtained. Reductions of equipment temperature of over 50 degrees Fahrenheit seem readily obtainable when specific attention is given to the arrangement and mounting of standard components. Such, and greater, gains will become common as soon as all concerned learn to focus on the concept of considering, calculating, testing for, specifying, and controlling equipment temperatures.

Discussion

Daniel Friedman (Naval Research Laboratory, Washington, D. C.): This paper discusses several very important factors which have not been given the consideration they deserve. The author shows that the specification of the ambient temperature is not sufficient to define completely the thermal environment under which the unit will operate. This is due to the fact that the temperature of the surrounding surfaces and the velocity of the air in the neighborhood of the equipment also will have an appreciable effect on its operating temperatures. He also points out that, with proper design, heat conduction could be used to reduce significantly equipment temperatures. This is particularly important at high altitudes where the air density is very low.

It should be noted that the paper presented concerns itself almost exclusively with self-cooled equipment and that the author's conclusions may not be directly applicable to blast or fan-cooled equipment. For example, in fan-cooled equipment the ambient air temperature is still the most important parameter.

While the author has presented a strong case for the revision or modification of the present testing and procurement specifications, it is felt that some of the proposals would not bring about the desired result. He seems to advocate that the specifications require that the surface or equipment temperatures, not the thermal environment, be controlled by specification. This places an unfair limitation on the equipment manufacturers. It would seem that the specifications should be revised to include:

1. The ambient temperature.
2. The temperature of the surrounding surfaces.
3. The velocity of the surrounding air.
4. The minimum life of the equipment.

This would supply a set of conditions for the equipment manufacturer, test chamber operator and designer, and airframe manufacturer to work toward. The specifications for equipment to be installed in high-performance aircraft should be based on the conditions to be expected in that type of aircraft and not on the conditions which existed in the relatively low-performance aircraft of a few years ago.

Airframe manufacturers, equipment manufacturers, and the military services are becoming more and more aware of the magnitude of the heat-transfer problems created by the high-performance aircraft being designed today. Although significant steps have been taken already, there is no doubt that greater progress toward the solving of these problems will be required in the near future.

Paul W. Franklin (Hatboro, Pa.): This writer wishes to compliment the author upon his concise presentation of the cooling problem of electric equipment in aircraft. It is interesting to note that some of the difficulties encountered are of administrative rather than of technical nature.

Regardless, if plain convection or forced cooling is used, air is required to transfer the heat rejected to the outside of the airplane. If no air is available, an artificial heat sink, like evaporation of liquids or transfer of heat to fuel, must be used since radiation alone is not sufficient. To cool equipment, one will attempt to continue to use the relatively simple ram air-cooling system as long as possible before changing to conditioned air or an artificial heat sink. In order to do this, the following steps should be considered:

1. As the author rightfully states, losses should be reduced to an absolute minimum.

2. The permissible hot spot temperatures of all components used should be raised to their maximum value, limited only by the characteristics of available insulation material. If class *H* insulation is used, permissible hot spot temperatures of at least 230 degrees centigrade are permissible for military applications; the writers experience indicates that even this value may be exceeded further, particularly for stationary components. However, there are certain exceptions, such as bearing grease which will not take much more than 150 degrees centigrade for limited periods and selenium rectifiers where the maximum permissible ambient at present does not exceed beyond 100–110 degrees centigrade. Rectifiers and other small sensitive components would have to be placed within the air conditioned area near the pilot.

3. A reduction of the probable life span of equipment exposed to high temperatures would have to be accepted. For instance, present conventional rotating electric equipment for aircraft is believed to have a life span of 1,000 hours under military conditions. Undoubtedly, this figure ultimately could be maintained for high-temperature equipment, but at an expense of weight, bulk, special cooling requirements, and reliability. Also, difficulties during mass production would have to be expected. Against this, the value of a simple, easily produced and replaced component should be considered, with a life span comparable to the period in between overhauls of a jet engine used on supersonic planes. Such a period would amount to approximately 500 hours. The life of insulation and the amount of permanent deformation (shrinking) is a function of the time-temperature integral; therefore 500 actual operating hours are equivalent to about 200–250 hours of actual full-load high-temperature operation, and it is not too difficult to build fairly reliable high-temperature equipment for such a duration. Of course, maintenance and supply would have to be adjusted to such a schedule.

4. One of the most important steps consists of the early recognition of the coolant consumption, preferably before the start of actual design work of the airplane. Unfortunately, this is not always the case, and very often electric equipment is squeezed into an existing limited coolant supply which is rarely adequate. This particular point is one of early co-ordination and administration and, in the writers opinion, much could be done to improve it.

Summarizing, this writer finds himself in full agreement with the present mode of thinking, to extend present cooling methods up to certain Mach numbers. To achieve this with the least amount of complication, systematically reduced life would have to be accepted and permissible hot spot temperatures would have to be raised as high as compatible with its characteristics, adjacent components, and the inherent functioning of the equipment (for example, excessive rise of copper losses due to increased resistance). Particular attention would have to be paid to a sufficiently large and expandable air supply, which should be incorporated into the basic design of the airplane.

Concerning some of the questions raised by the author, this writer would like to call his attention to some Government reports regarding high temperature problems as well as to the experience of manufacturers producing rotating electric equipment for aircraft.

Leonard J. Lyons: Mr. Friedman's objection to the control of surface temperature by specification is realistic, and the suggestion to use three of the most significant environmental factors would represent a very useful advance over current procedures. However, there should be no objection to specifying $T_a = T_w$, thus eliminating one of the factors, providing that the detail test procedure specifically requires that T_w be held equal to T_a, which is not unreasonable for qualification testing. At the same time, the test chamber size (relative to the equipment) and wall color (preferably dead black) can be specified, thus tying down several more important factors. It is assumed that air density, or altitude, would also be called out. The only significant factor then undefined is the orientation of the unit with respect to the cooling air flow, and this can be eliminated in qualification tests by requiring that the air velocity be low enough for natural convection to apply.

The specification will then only call out a temperature ($T_a = T_w$), an altitude, and a maximum allowable heat dissipation. Combined with known test conditions, this information would permit more adequate estimation of installation requirements than is currently possible. As pointed out in the discussion, fan-cooled units, or even lowered boxes, would still require further detailed consideration. It is most important to devise some method permitting rational estimates of derating necessary when the environment is hotter than specification and refrigeration is not feasible. Such estimates can only be based on an allowable value of component temperature or its equivalent.

Mr. Franklin's comments are also pertinent and useful. The concept of life being a function of the time-temperature integral is new to the author, and if applicable to components other than insulation, that is, a complete equipment assembly, could be particularly useful. With regard to the placing of more sensitive units in the air-conditioned areas, it may be found, with proper integration of the over-all airplane cooling problem, that other more efficient means are available. The use of cabin exhaust air is one possibility, and another is the bleeding of small quantities of refrigerated air from the cabin cooling package for cascade-type distribution through units arranged in the proper sequence.

Altitude Chamber Requirements for Evaluating Blast-Cooled Aircraft Generators

DALE H. SCOTT
MEMBER AIEE

ALTITUDE chambers have proved beneficial in the development of equipment to be installed in aircraft. By determining the conditions under which the equipment is to be operated and then simulating those conditions in altitude chambers, it has been possible to discover deficiencies and make corrections prior to installation of the equipment in aircraft. This procedure has made it possible to complete the development programs on the auxiliary equipment so that flight-test emphasis has been on the overall performance of the aircraft. Aircraft performance has been aided materially by the use of altitude chambers.

The altitude chambers are designed to simulate the conditions which are considered to be relevant to the satisfactory operation of the equipment undergoing investigation. In blast-cooled generators, it was early decided that minimum air-flow conditions determined the thermal ratings, and additional air flow due to higher speeds need not be simulated. With the advent of higher performance aircraft and the resulting aerodynamic heating, the limiting thermal conditions may occur at maximum rather than minimum speeds.

In this performance area, it will be necessary to simulate the conditions in altitude chambers to insure adequate generator cooling. As aircraft performance increases, blast cooling may be practical no longer, and other means of generator cooling may be required.[1,2] It is therefore necessary to review the problem and determine what steps can be taken to utilize existing altitude chambers to obtain necessary data, and what performance values will be required of future altitude chambers.

Description of Altitude Chambers

Altitude chambers designed to simulate blast-cooled-generator-flight conditions are composed of three systems: an air-preparation system, a test-cell system and an evacuation system. The air-preparation system delivers filtered air to the generator at the proper temperature after reducing the moisture content to some controlled or low level. The test-cell system houses the generator and includes temperature control equipment to maintain the test-cell space at the desired value. The evacuation system maintains the test-cell space at the desired pressure conditions. The feature of maintaining generator-inlet-air temperature at an entirely different value from the test-cell space requires additional equipment beyond that needed to simulate conditions for equipments which are self-cooled.

Types

RECIRCULATING

In the recirculating-type altitude chamber, the air is reconditioned repeatedly by the air-preparation system. The evacuation-system requirements are small, for it need carry only the system leakage plus sufficient fresh air to maintain the desired conditions. The air-drying requirements are small for the same reason. The rest of the air-preparation system becomes a difficult problem, for blowers to circulate large quantities of air at reduced pressure are inefficient, and a very large heat load must be removed by heat exchangers at reduced pressure. Most of the air-preparation system components must be located in the reduced pressure space, and the resulting reduction in rating increases the required size of these components. The blower motor can be located outside the low pressure space, but a shaft-seal problem is introduced. The low-pressure blowers require high-speed operation, and the resulting mechanical problems may be severe. In addition, the weight-flow rate of air to the generator must be measured over a large range of pressures.

Contaminants may be given off by the generator. In the recirculating-type chamber, these contaminants are difficult to eliminate. In certain altitude chamber tests, the presence of such contaminants have resulted in erroneous conclusions as to generator brush life.

NONRECIRCULATING

The air-preparation system of the nonrecirculating type is located external to the low-pressure space and delivers air to an expansion valve at or above atmospheric pressure. The heat-exchange problem is simplified, for the heat exchange is made at greater than sea-level pressure. The weight-flow rate of air to the generator also can be measured at greater than sea-level pressure.

The evacuation system of the nonrecirculating type is extensive for it must carry the entire cooling air in addition to the leak rate.

The test-cell system of both recirculating- and nonrecirculating-type chambers have separate circulating blowers, plus refrigeration and heating systems to maintain the test space at a temperature independent of the generator inlet air.

Existing Chambers

BASIS OF DESIGN PERFORMANCE

The cooling air is brought to an installed generator through a duct system. The entrance of the duct is located in an area such that during operation of the aircraft a pressure difference exists across the duct-generator system. This pressure difference on the earlier aircraft was caused by motion of the aircraft, or the propeller slipstream, or both.[3] In some cases an integral fan was used to aid ground cooling. As a result of analysis and experience, it was determined that minimum-flight speed should be the basis for 100-per-cent thermal rating. Increased speed furnished greater amounts of cooling air, consequently the cooling would be improved. Since it is reasonable to examine and simulate the worst conditions, only the minimum-speed conditions were considered in altitude-chamber design.

EVALUATION OF NAVAL RESEARCH LABORATORY NONRECIRCULATING CHAMBER

The nonrecirculating altitude chamber located at the Naval Research Laboratory was designed to simulate minimum aircraft-performance conditions for genera-

---

Paper 52-336, recommended by the AIEE Air Transportation Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Middle Eastern District Meeting, Toledo, Ohio, October 28-30, 1952. Manuscript submitted August 1, 1952; made available for printing September 5, 1952.

DALE H. SCOTT is with the Naval Research Laboratory, Washington, D. C.

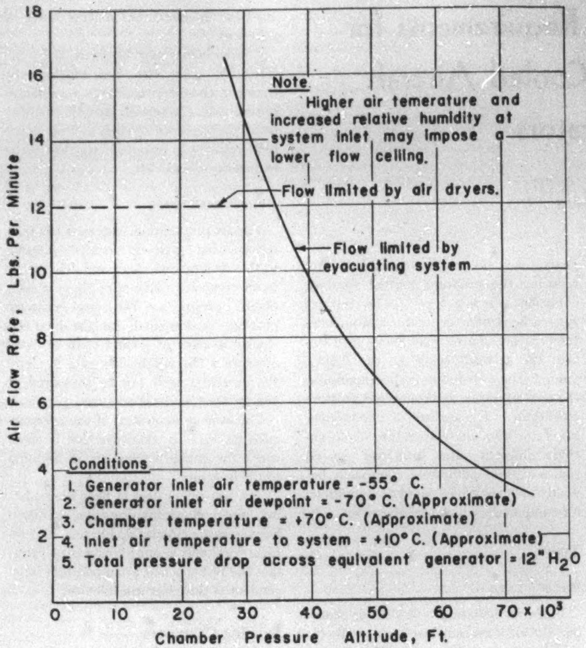


Figure 1. Typical flow characteristics of the Naval Research Laboratory nonrecirculating altitude chamber prior to the 1952 modification

tors 12 kva and smaller, from sea level to approximately 65,000 feet in altitude. As a result of increased generator ratings, the chamber performance limits were reviewed and the chamber was redesigned to increase its performance.

An experimental investigation of the chamber was made prior to modification to determine the maximum weight-flow rate of air for certain limiting conditions. The conditions included both 6 and 12 inches of water total pressure drop across the generator for a range of generator exit pressures (altitudes), while maintaining the generator inlet air at −55 degrees centigrade temperature and −70 degrees centigrade dewpoint. An adjustable restriction was used in place of a generator since it was desired to set a fixed value of pressure drop and independently adjust the flow to maximum.

As indicated in Figure 1, from approximately 27,000 feet to approximately 75,000 feet in altitude, the vacuum system determines the maximum flow for the conditions described. For continuous values of flow greater than approximately 12 pounds per minute, the air dryers will be damaged. The higher values of Figure 1 were held for a short interval to check the limits of the refrigerating system. Special blowers were substituted to obtain these high values, for the approximate upper limit of the regular blower was only 12 pounds per minute.

Conditions could not be maintained for altitudes greater than 75,000 feet, for the weight-flow rate of air was so small that the heat conducted into the ducting would not permit holding the inlet-air temperature at −55 degrees centigrade for greater altitudes.

Chamber Capacity Dependent on Conditions

The test requirements and weather conditions influence the operating limits. The refrigeration capacity of the air-preparation system is dependent on the weight-flow rate of air, the moisture content of the air as it enters the system, the air temperature as it enters the system, and the temperature of air required at the generator inlet. Referring to Figure 1, if the air which entered the air-preparation system had been +40 degrees centigrade instead of +10 degrees centigrade at the same relative humidity and all other test conditions the same, the refrigerating capacity might have limited the flow to somewhat less than the value shown. Conversely if a temperature of +10 degrees centigrade had been substituted for the −55 degrees centigrade at the generator entrance, it would have been possible to obtain data for altitudes greater than 75,000 feet.

Under test conditions in which the flow is limited by the evacuating system, the curve of Figure 1 will result for a large range in total pressure drop across an equivalent generator. The upper limit of pressure drop will be determined by the blower in the air-preparation system.

An increase in evacuation system capacity will displace the curve of Figure 1 to the right.

Review of Conditions to Be Simulated

Minimum Aircraft Performance Conditions Inadequate

In view of the higher operating performance of aircraft, the conditions to be simulated are subject to review. The rate at which heat is removed from a blast-cooled generator can be expressed by means of the following thermodynamic equation

$$Q = C_p W (T_o - T_i) \qquad (1)$$

where Q is the rate at which heat is removed, C_p is the specific heat of air, W is the weight-flow rate of air, T_o is the outlet-air temperature, and T_i is the inlet-air temperature. For a given generator design the following is assumed.

$$T_o - T_i = k(T_w - T_i) \qquad (2)$$

where k is a constant of proportionality and T_w is a limiting winding temperature. For a given design T_w may be set equal to a constant k_1, then by combining equations 1 and 2

$$Q = kC_p W(k_1 - T_i) \qquad (3)$$

It is evident from equation 3 that for a given generator design, the rate at which heat is removed by the cooling air is a function of the weight-flow rate of air and the inlet-air temperature. In the case of high-speed aircraft, the value of T_i may approach k_1 for some flight conditions, and on the basis of equation 3, no heat can then be removed by blast cooling regardless of the magnitude of W. Under those conditions, other cooling methods would be required.[1,2] For values of T_i

smaller than k_1, considerable heat can be removed if W is sufficiently large, and in this region altitude chamber studies are needed. Through basic engineering studies, it may also be possible to improve the heat-transfer coefficient which will result in a larger value of k in equation 3.[4]

ADDITIONAL HEAT PATHS

Aircraft generators normally are driven by the main engine through a suitable speed changer. The generators are usually cantilever mounted, and for thermal evaluation it is assumed that a negligible amount of heat is conducted through the attachment pad and drive shaft. Recent tests indicate that considerable heat is being conducted from the pad to the generator. The phenomenon was observed from flight-test data, test-stand data, and altitude-chamber data. Since the generator is not thermally insulated from the pad and drive shaft in most installations, additional data are needed as to their operating temperatures in order to simulate the conditions in altitude chambers.

DUCT AND GENERATOR LOSSES

Considerable variation in duct pressure losses occur on different installations. It is necessary, however, that all generators of a given rating be usable on all aircraft whose electric systems are designed to operate with generators of that rating. Generator and ducting airflow and pressure losses are controlled by specifications which have been prepared with the intent to provide this generator interchangeability.[5–7] These specifications are subject to review and clarification, the results of which will affect altitude chamber performance requirements.

Verification of Conditions to Be Simulated

FLIGHT TESTS

Very little flight testing of generator and blast-cooling installations are performed to determine if specification requirements are met, or if the conditions simulated in altitude chambers are encountered in flight as limiting conditions. This is largely because of the lack of suitable instrumentation to measure air flow in a blast-cooled installation during flight. Most methods for measuring air flow require a long straight run of ducting upstream from the instrumentation section, or the instrumentation disturbs the air flow excessively. On most fighter-type aircraft, the generator ducting is continuously curving.

In order to obtain reasonably accurate air-flow data in flight, a generator was calibrated in an altitude chamber for the flight conditions expected, and the mass rate of air flow was obtained for subsequent flight tests form the calibration charts.[8] The method is complicated but is a possible way of obtaining quantitative air-flow information in most aircraft installations. A means is thus provided for evaluating duct performance and for checking the validity of assumptions made regarding conditions to simulate in altitude chambers.

Modification of Existing Altitude Chambers

HIGH-SPEED FLIGHT

Altitude chambers designed to simulate only minimum aircraft performance conditions may require considerable modification in order to simulate high-speed conditions, particularly on generators whose ratings require maximum chamber capacity to simulate minimum performance conditions. It may be possible to obtain some high-speed performance data with no chamber modification by the use of low power-output generators.

In recirculating-type chambers high weight-flow rates of air can be furnished generators by the use of high-speed fans, such as superchargers, designed for aircraft use. These fans are inefficient, and thus will require powerful drive motors. Most of the drive-motor output will appear in the system in the form of heat, and extensive heat exchangers will be required to maintain the desired inlet-air temperature. Mechanical problems are introduced by use of high-speed fans, and the pressure seals on the high-speed shafts require extensive maintenance. In addition, these fans introduce high-level audio noise. If these problems can be solved, the recirculating type of chamber offers a means of obtaining high-speed flight data.

In nonrecirculating-type chambers, high-speed flight data should be obtainable by increasing the capacity of the evacuation system and the capacity of the blowers and air dryers of the air-preparation system. Increased refrigeration capacity should not be required. The modification of the Naval Research Laboratory nonrecirculating chamber is being performed on that basis. The capacity of the chamber is expected to be approximately 20 pounds of air per minute for a test-cell pressure equivalent to 50,000 feet in altitude and an air temperature of 40 degrees centigrade. By proper choice of generators, it should be possible to obtain considerable high-speed performance data with this equipment. These data are urgently needed for checking the validity of generator altitude rating charts, for use in preparing more complete generator performance specifications, and for future altitude chamber designs.

PAD TEMPERATURES

Means should be developed for maintaining the chamber drive-pad and drive-shaft temperatures at controlled values. Then upon determination of the representative aircraft pad and shaft temperatures, these values should be simulated during chamber valuation of generators. As an intermediate step, it may be necessary to thermally insulate the pads and drive shafts of altitude chambers and other generator drives used for generator thermal evaluation so that the tests will be performed under controlled conditions. Thermal evaluation performed on different types of generator drives will yield quite varied results, depending on the magnitude and direction of conducted heat.

Summary

The use of altitude chambers to simulate flight conditions for blast-cooled aircraft generators has materially aided the aircraft generator development program. These altitude chambers differ from those designed to simulate flight conditions for self-cooled equipment because of the need of a separate air-preparation system for forcing air through the generator.

In comparing the two chamber types, the recirculating type requires less evacuation and air-drying capacity, but contaminants are more difficult to remove from the air-preparation system and the heat exchange must be made at reduced pressure. The mechanical problems are aggravated on the recirculating type due to the high-speed blowers and the pressure seals on the high-speed shafts. In addition, high-speed blowers generate high-level audio noise.

Existing chambers were designed for simulating minimum aircraft performance. A review of requirements shows that high-speed performance also must be simulated. Conducted heat through the generator pad and drive shaft is quite significant. It is necessary to review and clarify specifications governing generator and ducting pressure losses so that realistic flight conditions will be simulated.

Increased flight testing is necessary to insure the validity of simulated conditions, and will be aided materially by adequate air-flow instrumentation. Means should be devised for simulating the accessory-pad and drive-shaft temperatures whenever generator thermal evaluations are performed.

Existing altitude chambers can be used for obtaining limited high-speed air-flow data and can be modified so as to extend their operating range. Generator high-speed air-flow data are urgently needed for generator-rating information and for future altitude chamber design requirements.

References

1. Cooling of Generator Equipment for High-Speed, High-Altitude Aircraft, **E. O. Naumann.** *Engineering Division Memorandum Report MCREXE-656-1898*, Air Materiel Command (Wright Field, Dayton, Ohio), September 1950.

2. Study of Aircraft Cooling Systems for Rotating Electric Equipment, **Cecil G. Martin.** *AIEE Transactions*, volume 71, part II, July 1952, pages 150–60.

3. Blast-Tube Cooling for Aircraft Generators, **Cyril G. Veinott.** *AIEE Transactions (Electrical Engineering)*, volume 63, July 1944, pages 520–25.

4. Heat Transfer Processes Within Rotating Electrical Equipment for Aircraft, **W. Robinson, J. R. Barnum, O. E. Buxton.** *AF 33 (038) 19842 Report Number 2*, Wright Air Development Center (Dayton, Ohio), October 1951.

5. Generators, 30 Volt D-C, Aircraft Engine Driven. *Military Specification MIL-G-6162* (Washington, D. C.), June 1950.

6. Generators and Regulators, Aircraft, Alternating Current. *Military Specification MIL-G-6099* (Washington, D. C.), April 1950.

7. Cooling Requirements of Power Plant Installations (Reciprocating Engines). *Bureau of Aeronautics Specification NAVAER E-59d* (Washington, D. C.), September 1946.

8. Measurement of Air Flow Through an Aircraft Generator During Flight, Part 1, **Joseph M. Marzolf.** *Report Number 3961*, Naval Research Laboratory (Washington, D. C.), April 1952.

Discussion

R. H. Simonds (Boeing Airplane Company, Seattle, Wash.): This is an extremely pertinent subject in that greater recourse to laboratory altitude testing is mandatory if the economical development of equipment capable of withstanding the severe environmental conditions imposed by installation in modern high-performance aircraft is to be kept abreast of the advances in aerodynamics and power plants. The author has presented a good review of the types of altitude chambers available and of the factors and limitations involved in the utilization of these chambers.

Considerable caution must be exercised in setting up simulated conditions for chamber testing if realistic results are to be obtained. One heat flow path that is frequently overlooked is radiation. In an airplane a piece of equipment generally is surrounded by other pieces of equipment operating at nearly the same temperature level. Therefore in the airplane there is little heat dissipated from the equipment by radiation. In setting up a test in the chamber frequently this is overlooked and the equipment is placed in the chamber with nothing to prevent radiation to the chamber walls. As equipment operating temperatures become higher this radiation factor will become increasingly important.

There is some doubt as to the possibility of mathematically deriving equations applicable to the cooling of rotating equipment without making oversimplifying assumptions. The air flow paths and velocities are so complex that it seems doubtful that a mathematical solution can be arrived at. For example, in equation 2 of the paper, k is assumed to be a constant. The results of our testing indicate that k varies considerably with altitude. Therefore the fact that inlet air temperatures may be lower at altitude does not mean that a lower mass flow of cooling air can be assumed. In fact, our results indicate quite the opposite. There is evidently a drastic change in the heat transfer coefficient as the altitude is varied. This is probably caused by the complex air flow pattern within the machine passages where the various flows may vary from laminar through turbulent flow and could conceivably be approaching sonic velocities in some of the passages.

Dale H. Scott: The author wishes to thank R. H. Simonds for his constructive comments and to agree with him that caution must be exercised in setting up simulated conditions for altitude chamber testing if realistic results are to be obtained.

Mr. Simonds has discussed the problem of radiation to the chamber walls. In the nonrecirculating chamber, realistic conditions can be simulated by maintaining the test-cell space at a value representative of the generator compartment temperature. Care must be taken to insure that the wall temperatures are essentially the same as the test-cell space. For some recirculating altitude chambers the problem of maintaining relatively high wall temperatures is difficult, and if the generator frame temperature is sufficiently high other means must be employed to prevent radiation.

The assumptions which resulted in equations 2 and 3 of the paper were introduced to furnish an approximation of the principles of blast cooling, and to illustrate limitations of the cooling method. The author concurs with Mr. Simonds that the value of k in the equations may not be constant as assumed. For a given mass flow, the effective coefficient of heat transfer probably will vary with rotor speed and with altitude. Experimental means must be employed to obtain actual data on this complex heat transfer problem. The limiting thermal conditions may exist at high-speed flight conditions rather than during minimum speed flight, and altitude chambers provide a good means for obtaining necessary data.

Feeder-Fault Protection for D-C Aircraft Generators

WILLIAM W. DAVIS
ASSOCIATE MEMBER AIEE

THE electric system of the modern airplane must provide continuity of electric power to all load distribution circuits regardless of the occurrence of faults. The aircraft industry has widely accepted the necessity of generator feeder-fault protection to accomplish this objective. There is, however, no general agreement to be found when determining the means of obtaining this protection. Almost as many methods are now in use as there are aircraft manufacturers, and most of these methods operate to a degree of satisfaction.

This paper reviews the more widely used methods of obtaining protection and discusses the problem of feeder-fault protection for d-c aircraft generators.

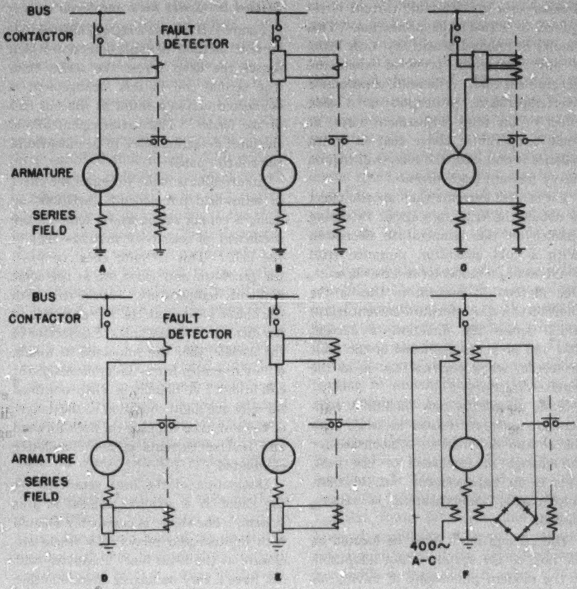
For the purpose of the paper, the generator feeder is defined as the power leads from the generator to the main bus. Protection of the main bus and of the distribution system are equally important but are beyond the scope of this paper.

The Need for Feeder-Fault Protection

Instances of fault conditions have been recorded since the first electric system was installed in an airplane. The probability of faults occurring and the severity of faults have increased with each increase

Paper 52-332, recommended by the AIEE Air Transportation Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Middle Eastern District Meeting, Toledo, Ohio, October 28–30, 1952. Manuscript submitted February 5, 1952; made available for printing September 8, 1952.

William W. Davis is with the Boeing Airplane Company, Seattle, Wash.



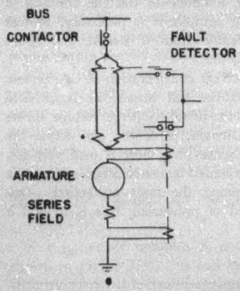


Figure 1. Methods of fault detection

in airplane size, electric system complexity, and generator capacity.

The generator feeders are particularly vulnerable because of airplane design. Weight economy has dictated the use of the airplane structure as a ground return. Hence a short circuit to structure, as well as to other circuits, constitutes a fault condition. The feeders are long, extending from the generators in the engine nacelles to the main power bus which is, for most modern designs, centrally located in the fuselage. Normal airplane vibration subjects these leads to chafing and abrasion and during combat they are exposed to gun fire. They are also subjected to engine vibration which, in some airplanes, is extremely severe. Most open-circuit conditions, as well as short circuits, eventually become ground faults because of airplane vibration.

Almost all present-day multiengine airplanes have an electric system in which two to six d-c generators are operated in parallel. This represents an installed generating capacity of 6 to 72 kw. Under fault conditions these systems are capable of maintaining fault currents of up to 8,000 amperes. Tests and experience with systems of these capacities have shown that the "burning clear" of faults is not reliable or effective protection. Burning clear has proved to be:

1. Uncertain—instead of burning clear, the wire may weld and stall the system.
2. Too slow—irreparable damage may be done before the fault clears.
3. Not permanent—airplane vibration may cause intermittent fault with resultant arcing throughout flight.
4. Fire and explosion hazard.
5. A cause of structural damage.

The dependence of the airplane on continuity of electric power and the destructive nature of feeder faults make it essential that feeder-fault protection be provided.

Protective History

The only generator protection provided for many years was the reverse-current relay. This relay, still an integral component of present systems, operated when current flow at the contactor was in the direction opposite to normal. This provided an effective means of removing feeder faults from the bus as long as the relay interrupting capacity was not exceeded. However, no means was provided to de-energize the field of the generator, so the generator continued to feed the fault. Thus dependence was still placed on the fault to burn clear and the danger of fire and structural damage was not removed. In addition, the reverse-current relay was not trip-free. Intermittent faults could permit the contactor to reclose, often resulting in damage to the contactor and increasing fault damage due to power from the bus.

Fusible limiters, to act as backup protection for the contactor, were next used. These were designed to blow at above normal overload levels and provided some protection on bus and distribution system faults as well as on feeder faults.

Another solution was the use of manually reset reverse-current circuit breakers located between the generator contactor and the bus. This acted as backup protection for the reverse-current relay and incorporated an auxiliary set of contacts to open the field circuit when the circuit breaker tripped. Lack of time delay in the early circuit breakers allowed them to trip on normal system reverse-current surges and their location in the nacelles adjacent to the generators made them inaccessible for resetting. The result was that for some time these circuit breakers caused far more system troubles than they provided protection. In later designs, addition of inverse-time delay co-ordinated with the reverse-current relay has minimized the danger of nuisance tripping, and the modern tendency toward use of a centralized bus permits access for resetting. Some of the recent circuit breakers are provided with an electrical reset to allow remote operation.

The reverse-current circuit breaker and the fusible limiters operated only on faults greater than normal overload levels. The danger of fire, while not as great, is still present on lower-current faults.

The foregoing circuit components provided detection of feeder faults by reverse-current flow from the bus through the contactor. No fault protection was provided if a feeder-fault occurred when the generator was unable to go on the bus because of low voltage, faulty contactor, open differential coil, or reversed generator polarity. The probability of a fault is equally great with the generator on or off the bus.

Modern Methods of Feeder-Fault Protection

A great many systems have been devised to overcome the shortcomings of the early feeder-fault protection methods. Almost all that have proved satisfactory and found wide usage are of the balanced-current or differential-current types or a combination of the two. The balanced-current method requires two parallel identical feeders for each circuit to be protected. A fault on either cable causes a current unbalance in the two cables. The unbalance is used to trip a protective device at each end of the feeder which causes the field relay to trip and the contactor to open. The differential-current method compares the current in the two ends of a circuit, and when they differ by more than a fixed amount the protective device is actuated.

The major applications of these two fault protection methods will be discussed in detail.

FAULT DETECTION

Method A—Series Field and Positive-Feeder Coil

Figure 1(A) illustrates one method of differential-current fault protection. The fault-sensing relay is polarized and has two coils wound on a common magnetic circuit. One coil consists of a single turn of heavy copper and is connected in series with the feeder at the bus end of the feeder. The other coil consists of many turns of small wire and is connected across the generator series field.

One serious drawback to this and all other systems where series field sensing is used is the effect of the extreme temperature range of the series field. The voltage drop sensed by the coil across the field can vary by a factor of 2, while the series coil senses a constant ampere-turn value for any given load. Thus a perfect balance between the two coils can be obtained for only one generator temperature. The balance is then set at the highest allowable generator field temperature. This determines the minimum fault current at which the relay will trip and eliminates the possibility of nuisance trips on forward currents. For lower field temperatures the field voltage drop is smaller so a greater fault current is required to actuate the protection. This results in reduced sensitivity with little, if any, protection provided when the generator is cold. A measure of generator overtemperature protection is a side effect of the relay adjustment since, at some temperature above that at which balance is obtained, the relay will trip on heavy forward load currents.

On reverse currents the opposite effect is noted, the tripping current becoming smaller as the temperature decreases. With a cold generator, nuisance trips could occur on moderate reverse currents. One method of preventing this is the addition of a relay (forward-current relay) which senses the direction of current flow and prevents operation of the fault protection when current flow is in the reverse direction.

Some generators now include a temperature-sensitive resistor in series with the *D* lead designed to compensate for the change in resistance of the field. This is an improvement but, at best, only partial compensation is accomplished.

The sensing relay must be located at the bus, so the system is not adaptable to the modern philosophy of having all control and protective devices centrally located in an easily replaceable control panel. Use of a polarized relay provides a measure of fail-safe protection. Loss of the long lead to the series field would decrease sensitivity, but the fault protection would still operate on reverse current from the bus. Loss of this lead, with either polarized or nonpolarized relays, would result in loss of all protection if a fault occurred when the generator was not connected to the bus.

Method B—Series Field and Shunt

Figure 1(B) shows a second differential-current method of fault detection which senses the drop across the series field. The second coil in this arrangement is connected across a shunt at the bus end of the feeder. This arrangement allows the fault detection relay to be included in the control panel.

An attempt is made to cancel the effect of series-field temperature variations by using a copper shunt whose temperature coefficient of resistivity matches that of the field. This is quite effective when the generator and shunt are at the same ambient temperature. However, with the shunt located at the power center in the airplane fuselage and the generator in the nacelle, they are subjected to widely different temperatures under normal flight conditions. A balance is then obtained for only one flight condition so the danger of the nuisance tripping on both forward and reverse currents is not completely eliminated.

Dissipation of the heat generated by the shunt is a serious problem in this system. The shunt is commonly located with limiters and relays in a shield enclosure at the main bus. Excessive heat can have a very damaging effect on other circuit components and the addition of cooling equipment is necessary, though undesirable from a weight and space standpoint.

A further consideration in this system is that the coil across the series field senses the inductive surge voltage during load changes, while the coil across the shunt senses only normal load voltages. Surges caused by load changes are capable of tripping the fault protection. One method of preventing this is to add a

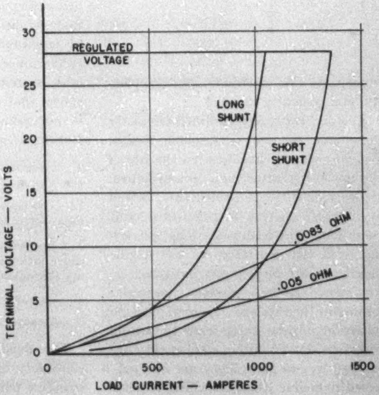


Figure 2. Generator external characteristic

reactor in series with the fault-sensing coil which is across the series field. The inductance of the reactor is adjusted to absorb the series-field voltage surge. A second method of eliminating this is to add a surge override relay. This relay operates on the surge and takes the ground-fault protection out of the circuit for a specified time delay. This adds a time delay co-ordination problem as well as reliance on an extra relay and, during normal fault conditions, adds to the time required to remove the fault from the system. This system is considered to be fail-safe. Loss of a lead to the series-field sensing coil would again result only in lowered sensitivity of the relay when the generator is connected to the bus and would result in loss of protection when the generator is not connected to the bus. Loss of a lead to the shunt would result in a false trip when the load reached the minimum trip setting with the generator connected to the bus and would have no effect with the system off the bus.

In some applications of both methods A and B it has been found necessary to return the ground lead of the series-field sensing coil back to ground at the generator instead of to structure at the control panel. It was found that a potential sufficient to affect fault protection calibration existed between airplane ground points during heavily loaded conditions.

Figure 1(C) illustrates a variation of method B. The two coils across the shunt are cumulatively wound and their total flux is balanced against the flux of the coil which senses across the series field. Operation is similar to that of method B.

Method D—Negative-Lead Shunt and Series Coil

A variation of methods A and B is shown in Figure 1(D). A coil in series with the feeder is balanced against a coil which senses the voltage drop across a shunt in the generator negative lead. This eliminates the possibility of nuisance trips caused by variations in generator series-field temperature. Shunt cooling is not as great a problem since heat generated by the shunt would not be as serious in the nacelle. Again the fault-protection system could not be installed in a centrally located control panel.

Method E—Two Shunts

The 2-shunt method of differential-current protection is shown in Figure 1(E). The balanced coils sense the voltage drop across two identical shunts, one in the generator negative lead, and the other at the bus end of the feeder. The shunts are made of Advance metal or some other material with a low temperature coefficient of resistivity so the possibility of nuisance trips due to shunt temperature changes is eliminated. The shunts are purely resistive so there is no possibility of nuisance trips from forward- or reverse-current surges. The protection level is relatively constant for all load and generator speed conditions and can be adjusted to values below 100 amperes provided the shunt and control lead resistances are held within reasonable tolerances.

Whenever shunts are used there is the possibility of loose- or high-resistance terminal connections, which would result in loss of sensitivity or nuisance tripping. The shunts also introduce a serious heat dissipation problem as discussed earlier. Sensing is accomplished with a single relay which can be conveniently located in a control panel installation.

Both polarized and nonpolarized sensing coils have been used with this type of installation. The polarized relays improve the fail-safe features, since loss of negative shunt sensing results only in lowered sensitivity whereas a nuisance trip would result if the same lead were lost with a nonpolarized relay. With either type of relay, loss of a lead to the positive shunt would result in a nuisance trip when the load current reached the minimum trip setting.

Method F—A-C Reactor

Figure 1(F) illustrates a system which introduces an a-c signal to sense current unbalances. The fault detectors act as saturable reactors and consist of doughnut-shaped laminated cores with primary and secondary windings. The generator feeder wires pass through the openings in the center of the cores. The primary windings are connected in series with 400-cycle a-c supply. The secondaries are connected series opposed so that there will be no net secondary voltage for equal flow of load current through the generator feeder cables. A d-c current unbalance produces a difference of primary impedance with consequent difference of secondary voltages. The net difference is rectified by a bridge circuit and, upon reaching the preset trip value, operates the fault-protection relay.

The relay can be adjusted to operate on a relatively small value of fault current and, like the 2-shunt method, is free of nuisance tripping from current surges and temperature variations. There is no d-c power loss with the resulting heat dissipation problems.

The system requires approximately 5 volt-amperes of a-c power for normal operation. It is claimed that loss of a-c excitation does not result in complete loss of fault detection but a greater value of fault current is required to trip. This would mean, of course, that the fault detection must operate during the transient period. Except for reactors the unit can be adapted to a control-panel installation.

The principal disadvantages are reliance on a-c power for d-c protection, the additional relay and rectifiers required, and the difficulty of manufacturing reactors which will have matched characteristics over a wide range of direct current.

A measure of balanced-current feeder protection can be obtained in a 2-wire feeder by installing a detector in one end of one wire and another in the opposite end of the second wire. The protection level would be a function of the location of the fault with a possibility of no protection at or near the unprotected end of each feeder wire. This arrangement also can be used in conjunction with the reactors in both generator leads to give more complete protection.

Method G—Balanced Current and Differential Current

A method of feeder-fault protection which combines the balanced-current and differential-current methods of fault detection is shown on Figure 1(G). A differential-current relay, mounted at the generator with coils in series with ground and feeder leads provides protection against generator and generator-lead faults. Balanced-current coils wound on the same relay protect against faults in the generator end of the feeder. A balanced-current relay installed at the bus end of the feeder completes the feeder protection.

A dual-channel feeder is required. The system is fast-operating and is not subject to nuisance trips on current surges or temperature variations. The protection level is a function of the fault location, with the minimum trip value occurring for faults at either end of the feeders and the maximum value for faults at the middle of the feeder.

With one relay at the bus and one at the generator, this fault detection system is not adaptable to the control panel type of installation. The balanced-current relays sense high resistance junctions in the feeder as well as feeder faults.

ISOLATION OF FAULT

In all of the fault-protection systems covered, actuation of the fault detection merely closes the field relay trip coil circuit. Thus, even with reliable fault detection, it is necessary to have a voltage source to trip the field relay and the contactor must have sufficient capacity to interrupt all current fed to the fault by the bus.

Field Relay

Dependence of the fault protection on the tripping of the field relay is one of the most serious drawbacks of present systems. Field relay design has progressed to the point where as low as 4 volts may be sufficient for tripping, yet on extreme faults the system voltage can be depressed below this value. Thus, even with perfect fault detection, the system can be stalled and voltage to essential loads lost.

Field relay tripping voltage is maintained for much lower values of fault resistance with a short-shunt generator than with a long-shunt machine. Typical external characteristic curves of a 400-ampere 28-volt d-c generator are reproduced on Figure 2. This shows that a fault resistance of 0.0083 ohm depresses the generator voltage to 4 volts with the long-shunt connection. When the machine is connected short-shunt the same fault resistance depresses the voltage to only 8.6 volts. A fault resistance of 0.005 ohm then is required to depress the voltage to 4 volts. Improved feeder-fault protection then is provided by using short-shunt generators. Use of short-shunt machines is also desirable to aid in clearing bus and distribution system faults since, as is also shown in Figure 2, a much greater fault current is sustained for a given fault resistance. A fault current of 480 amperes is delivered by the long-shunt generator with an 0.0083-ohm fault resistance, whereas the same generator connected short-shunt is capable of a 1,030-ampere fault current.

Many systems return the negative of the field relay trip coil to the generator *D* lead in order to utilize the drop across the series field as additional tripping voltage during severe faults. Tests have shown that with present self-excited generators the output current reduces to a very low value when the voltage is depressed by an extreme fault. The series field drop is then small so that there is only a very narrow range of fault resistances over which increased protection is provided. This would be improved materially if the generators could be designed to put out substantial current (at least half rated) under short-circuits, zero-voltage conditions. Connecting the field relay trip coil to the *D* terminal of the generator is undesirable because the wire is long and is subjected to engine vibration. If the lead were lost, all feeder fault and overvoltage protection would be lost in some systems. In other systems, loss of this lead would decrease the reliability of the ground-fault protection.

Another solution is the use of a changeover system which transfers the field relay trip coil to an alternate source of power when the voltage is depressed below a safe value. This is an operable solution and has found some use but it places dependence on a separate voltage source and on the operation of the changeover relay. Also, it is considered questionable design practice to have two separate voltage sources on adjacent contacts of the same relay.

The only real solution to this problem is to design the fault protective devices so that they utilize fault power to remove the fault. One system not previously discussed which does this is identical in detection to that shown in Figure 1(A). However the fault-detection coils are the trip coil of the field relay. This particular system has all the disadvantages of method A plus the further disadvantage of requiring a high value of fault current to trip the latch-type field relay. A second possibility would be to use the 2-shunt configuration of method E, again having the differential coils be the trip coil of the field relay. The main problem is, of course, to obtain sufficient power to trip the latch-type relays without sacrificing fault-detection sensitivity. This has restricted use of these methods of detection but it is felt that this is the right approach to the problem.

Backup Protection

The final consideration of a feeder protection system is the ability of the contactor to interrupt the current fed to the fault by the bus. Numerous instances of frozen or welded contacts or contacts which have burned away have been recorded and used as evidence to support the claim of need for contactor backup protection.

Tests and experience have shown that most contactor failures are a direct result of chattering during fault conditions. When feeder-fault protection is used this can be caused only by faults on or beyond the bus. On this type of fault the bus depresses the voltage to a low enough value so that the contactor drops out interrupting the current fed to the bus by the associated generator. With the fault removed, the generator voltage immediately builds up, reconnects to the bus, is depressed, and the cycle repeats, often until the contactor is destroyed. As a result of varying fault power, distribution system limiter clearing time is greatly extended, or the limiters do not clear at all.

Two types of backup protection are presently in use, the reverse-current circuit breaker and double contactors. Neither provide any protection for the type of fault just described where backup is most needed. Both are intended to increase the interrupting capacity of the system. The reverse-current circuit breaker performs this function on high-current feeder faults and with auxiliary contacts also opens the field circuit. It has the important advantage that it is operated by fault power. The circuit breaker adds an airplane weight penalty and presents a co-ordination problem with the reverse-current relay. The double contactor also increases interrupting capacity on feeder faults, overvoltage conditions, and extreme reverse currents. It has the advantage that it is automatically reset if the generator is able to supply power to the system.

The need for backup could be reduced, if not eliminated, by incorporating in the contactor the current interrupting capabilities of the reverse-current circuit breaker. To do this no doubt would require use of a latch-type contactor. This is desirable because it would at the same time eliminate the chatter problem of bus and distribution system faults.

Protective Device Tolerances

Fault protective systems should operate at a low value of feeder-fault current yet allow extreme currents to be fed to the bus during through-fault conditions without nuisance tripping. This requirement makes the allowable manufacturing tolerances of protection components very critical.

For example, in the 2-shunt system it is common to allow a shunt tolerance of ±2 per cent. If one shunt had the maximum plus tolerance and the other the maximum minus tolerance, a 16-ampere unbalance would be sensed by the coils at 400 amperes load and 60 amperes unbalance at a load of 1,500 amperes. Forward surge currents of this magnitude have been measured during bus and distribution system fault tests. Even if the balance of the coils and of the leads were perfect the minimum trip setting of the fault protection must be above 60 amperes to prevent nuisance tripping on through-fault or current surge conditions. Any unbalance in sensing-lead resistance would increase the load unbalance seen by the sensing coils, and the unbalance of the coils themselves must be included also. As a result, a tripping value much below 100 amperes cannot be achieved without the possibility of nuisance tripping. The same type of unbalances would apply to the a-c reactor system, and to a lesser degree to the other differential-current and balanced-current systems. Shunts are now made which can meet the ±2 per cent requirement but it is questionable whether this close a balance could be held with saturable reactors. Relays can be balanced within fairly close limits. It is of course important that the coils of the relay be designed and installed so that they are always at the same temperature.

All the systems which depend on long control wires to transmit the sensing voltage to the control panel or another remote point could be improved by designing the coils with sufficient resistance so that a variation in lead wire resistance would have little effect on calibration. It is difficult to obtain a good resistance balance of relay leads since normally the leads to one coil are long and go through a number of junctions and the leads to the other coil are relatively short. Most differential-relay sensing coils have a resistance of the order of 2 to 3 ohms and a variation of 0.1 ohm in lead resistance has a considerable effect on the tripping values. If these could be increased to resistances of the order of 10 to 20 ohms the effect of the lead resistance would be minimized and system sensitivity and reliability improved.

Review

The systems covered constitute the principle feeder-fault protection methods now in aircraft use. Each of the systems has merit but not one is free of major criticism. Some require an additional power source for proper operation, others introduce additional problems such as equipment cooling, and still others require additional relays, reactors, rectifiers, or other circuit components to insure that the system performs its designed function under all operating conditions. The system thus becomes complex, and as was the case with the early reverse-current circuit breakers, the protection systems themselves may present more problems and initiate more troubles than those they were intended to protect against. The 2-shunt method (method E) and method G, which combines differential-current and balanced current protection, appear to introduce the fewest problems and to be the most trouble-free of the systems considered.

A summary of the factors in favor of each of the systems covered and of the items where further design and development work is needed suggests the following list. A feeder-fault protection system should:

1. Protect the generator as well as the feeder.
2. Utilize fault power to clear faults.
3. Be independent of other power sources for fault sensing.
4. Require low power for operation.
5. Provide a constant protection level of the order of 100 amperes.
6. Use a minimum number of relays and circuit components to obtain protection.
7. Require no additional components to protect or insure proper operation of the protective devices.
8. Introduce no additional system problems such as equipment cooling.
9. Be completely free of nuisance tripping.
10. Be fail-safe yet permit generator to provide power to the bus as long as the system is able.
11. Be compact and light weight.
12. Be adaptable to a control-panel type of installation.

Future design and development work should be directed toward achievement of a system incorporating the twelve foregoing items with particular effort being directed toward system simplification and toward utilization of fault power for fault isolation and generator deactivation.

References

1. Electric-Circuit Fault-Protective Principles as Applied to D-C Aircraft Systems, **R. H. Kaufmann.** *AIEE Transactions*, volume 63, 1944, pages 333–44.

2. Protection of the Airplane Main Bus, **D. W. Exner.** *AIEE Transactions*, volume 68, part II, 1949. pages 979–82.

3. Control and Protection of Aircraft D-C Power Systems, **B. O. Austin.** *AIEE Transactions (Electrical Engineering)*, volume 65, November 1946, pages 741–5.

4. D-C Power Systems for Aircraft—Parts VI and VII, **R. H. Kaufmann, H. J. Finison.** *General Electric Review* (Schenectady, N. Y.), June 1946, pages 46–52, and August 1946, pages 41–48.

No Discussion

An Approximate Transfer Function for the Analysis and Design of Pulsed Servos

R. G. BROWN
ASSOCIATE MEMBER AIEE

G. J. MURPHY
ASSOCIATE MEMBER AIEE

MUCH has been written concerning the use of frequency response techniques in the design and analysis of linear control systems with continuous error detection and control. Application of these techniques to the problems of discontinuous control system design, however, has not been treated extensively. The purpose of this paper is to present an approach to the analysis and synthesis of a particular kind of discontinuous control system, the pulsed servo, based on modified frequency response methods.

Description of Pulsed Servo Operation

Discontinuous control systems are those which employ intermittent error detection or periodic application of the control effort. The pulsed or sampling servo is a particular case of the discontinuous control system in which the system input and output are continuously varying functions of time, while the error detection is done only at discrete intervals. This results in a discontinuous error signal in the form of a sequence of impulses of varying magnitudes.

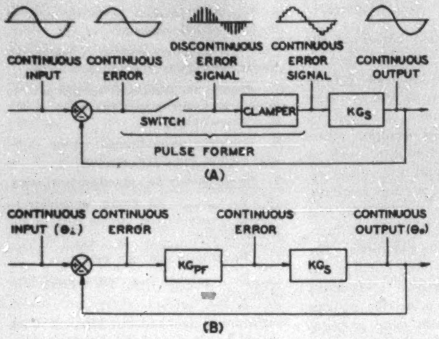
Figure 1(A) shows a block diagram of a pulsed servo system with the usual clamping or holding action introduced after the error detector. The effect of clamping is to convert the discontinuous error signal to a piecewise continuous function of time by providing an output signal which, during any given sampling interval, is proportional in magnitude to the error impulse at the beginning of that interval. As the sampling interval approaches zero,

---

Paper 52-314, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Middle Eastern District Meeting, Toledo, Ohio, October 28–30, 1952. Manuscript submitted August 1, 1952; made available for printing September 5, 1952.

R. G. Brown and G. J. Murphy are with the General Motors Corporation, Milwaukee, Wis.

The authors wish to acknowledge the assistance of J. Stuber in the preparation of drawings and manuscript for this paper.



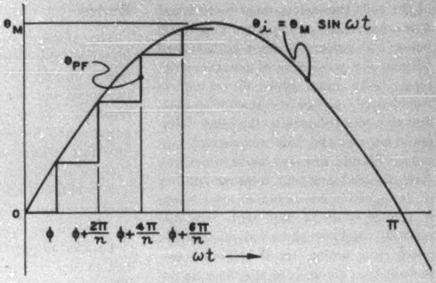


Figure 1 (left). Functional block diagrams

A. Pulsed servo system with clamper
B. Equivalent system with pulse former transfer function

Figure 3 (above). Initial portion of pulse former input and output

the response of the pulsed system with clamping approaches that of the corresponding continuous data system. For increased sampling periods, however, the data supplied to the controller near the end of any sampling period become "stale" with the result that the response of the pulsed system departs considerably from that of the continuous data system.

A number of methods for the analysis of systems such as the one just described have been developed in the past. The method of Hazen,[1] for example, involves a step-by-step solution of the system using the differentiating operator and the matching of final and initial conditions from one sampling interval to the next. The methods of Hurewicz[2] and of Smith and Lawden[3] employ time sequences rather than continuous functions. Experience with the use of these methods reveals that, although all three are completely rigorous, they involve lengthy and tedious manipulation in all but the simplest of cases. Furthermore, these methods do not seem to be readily usable in synthesis of complicated systems because of the relationships between input and output functions of the pulse modulators when expressed in terms of sequences or differential operators.

A fourth method of analyzing systems of the type considered here is that given by Linvill.[4] This method, although it is based on a rigorous foundation, makes possible the use of approximations leading to the drawing of a "$\Sigma K\beta$" locus in the KG plane. When the ratio of input frequency to switching frequency is large, this locus can be used satisfactorily in the same manner as the familiar Nyquist diagram. In fact, it is this method of analysis of pulsed systems with clamping which suggests that a very simple approximate transfer function may exist for the impulse modulator-clamper combination. The derivation of such a transfer function is the subject of the following section.

Derivation of the Approximate Transfer Function of the Pulse Former

In deriving an expression for the approximate transfer function of the pulse former, it can be assumed that all harmonic components of the pulse former output except the fundamental are negligible. The higher harmonics generated by the pulsing operation are shown to be relatively small in Appendix III. It will be seen in Figure 1(B) that the desired transfer function KG_{PF} is dependent primarily upon the ratio $n = \Omega/\omega$ where

$\omega = 2\pi/T_R$ = angular response frequency
T_R = response period
$\Omega = 2\pi/T_s$ = angular switching frequency
T_s = switching interval or sampling period

Figure 2 shows the relationship between the clamper output θ_{PF} and its Fourier fundamental θ_1 for a case where n is an integer (those instances where n is not an integer are discussed in Appendix II) and the error sampling begins at $t=0$. A more general case is presented in Figure 3 where the input wave θ_i leads the sampling by an angle θ.

For any integer value of n, the expressions for the coefficients of the cosine component and sine component of the Fourier fundamental of the output wave θ_{PF} are

$$A_1 = \frac{1}{\pi}\int_{\phi}^{\phi+2\pi} \theta_{PF} \cos \omega t \, d\,\omega t$$

and

$$B_1 = \frac{1}{\pi}\int_{\phi}^{\phi+2\pi} \theta_{PF} \sin \omega t \, d\,\omega t$$

The output of the clamper can be expressed as

$$\theta_{PF} = \theta_M \sin\left\{\phi + \left[\frac{2\pi(m-1)}{n}\right]\right\}$$

where

$$\phi + \frac{2\pi(m-1)}{n} < \omega t \leq \phi + \frac{2\pi m}{n}$$

After performing the indicated substitution for θ_{PF}, the expressions for A_1 and B_1, as shown in Appendix I, reduce to

$$A_1 = -\frac{\theta_M n}{\pi} \sin^2\left(\frac{n}{\pi}\right), \quad n = 3, 4, 5, 6 \ldots$$

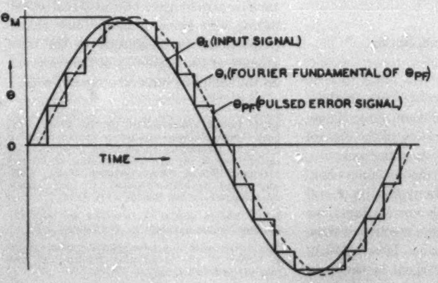


Figure 2. Relationship between input to pulse former, output of pulse former and equivalent Fourier fundamental

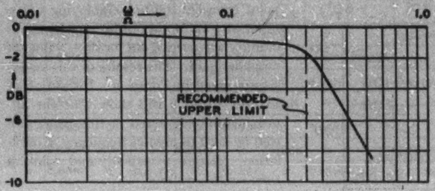


Figure 4 (above). Gain characteristics of pulse former

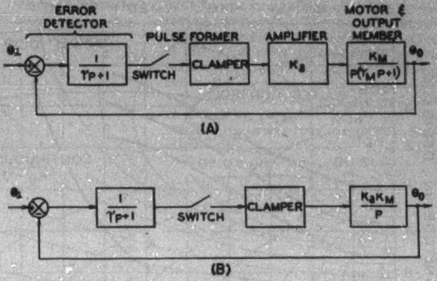


Figure 6 (right). Functional block diagrams of illustrative system

A. Block diagram of a typical pulsed servo

B. Reduced diagram assuming $\tau_m << \tau$

and

$$B_1 = \frac{\theta_M n}{2\pi} \sin\left(\frac{2\pi}{n}\right), \quad n = 3, 4, 5, 6 \ldots$$

Hence, the total fundamental component of the pulse former output is given by

$$C_1 = \frac{\theta_M n}{\pi} \sqrt{\sin^4\left(\frac{\pi}{n}\right) + \frac{1}{4}\sin^2\left(\frac{2\pi}{n}\right)}, \quad n = 3, 4, 5, 6 \ldots$$

and its phase angle is

$$\phi_1 = \tan^{-1} \frac{-2\sin^2\left(\frac{\pi}{n}\right)}{\sin\left(\frac{2\pi}{n}\right)}, \quad n = 3, 4, 5, 6 \ldots$$

Thus it becomes possible to account for the dynamics of switching or pulsing by appropriate substitution of the pulse former approximate transfer function described by the magnitude and phase characteristics of Figures 4 and 5. In general, the net effect of pulsing a servo is to introduce into the system a simple phase lag, with small magnitude change, over the bandwidth normally of interest.

The transfer characteristics in Figures 4 and 5 are universal in that they can be applied to any device which functions as a switch-and-clamper combination. These curves make extremely simple and conventional the determination of the feedback transfer function of a given continuous data system to which a pulse former has been added.

Analysis of a Typical Pulsed Servo

The example outlined in some detail in this section illustrates the manner in which the approximate transfer function can be used to analyze or predict the performance of a typical sampling servomechanism. Actually, the basic approach which is followed is of particular value when the problem is that of system design rather than analysis.

It should be noted that the complexity of the continuous portion of the system does not materially increase the work required to arrive at a satisfactory solution because of the assumed independence of the pulse former.

The block diagram of Figure 6(A) shows a typical pulsed servo, the continuous portion of which includes an error detector with an inherent delay τ, a linear amplifier with a gain constant of K_a, and a motor having a gain constant of K_m and an inherent delay τ_M. To simplify the analysis, let it be assumed that $\tau_M << \tau$ so that the system block diagram reduces to that of Figure 6(B). As a result of system mechanization, the following parameters are known or given: $\tau = 0.33$ second, $K_m = 0.50$ radian per volt second, $T_S = 2$ seconds, K_a = variable.

To analyze the pulsed servo, it is first necessary to plot the familiar magnitude and phase versus frequency curves for the continuous portion of the open loop transfer function. These curves, shown in Figure 7, are asymptotic approximations based on an assumed amplifier gain of unity. The effective magnitude change

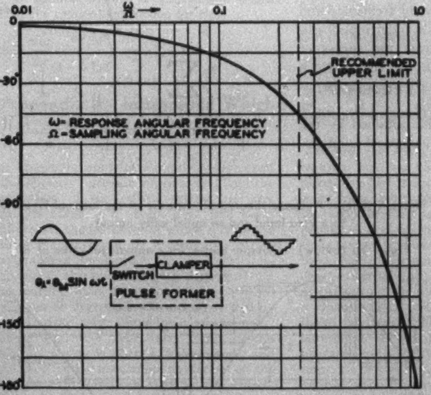


Figure 5. Phase characteristics of pulse former

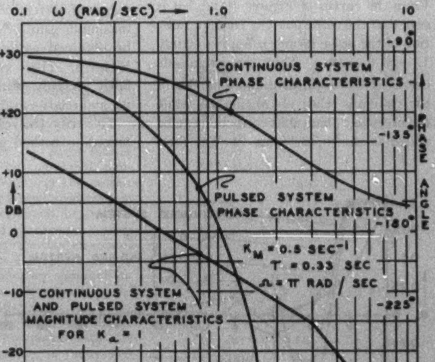


Figure 7. Open loop frequency response characteristics of the loop of Figure 6

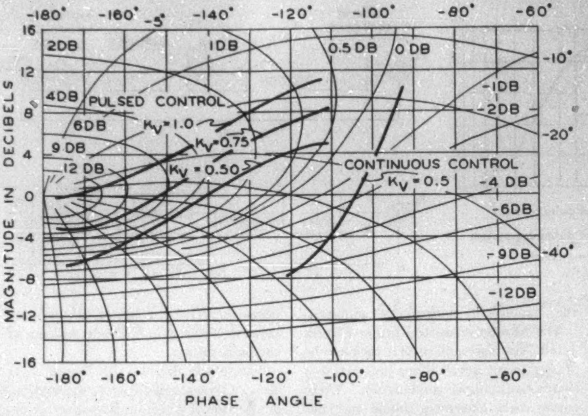


Figure 8. Closed loop characteristics of the loop of Figure 6

and phase shift due to the introduction of discontinuous control action may be determined at discrete frequencies in the following manner:

For

$\omega=0.5$ radian per second

and

$\Omega=\frac{2\pi}{T_s}=\frac{2\pi}{2}=3.14$ radians per second

the value of n is

$n=\frac{\Omega}{\omega}=6.28$

or

$\frac{1}{n}=\frac{\omega}{\Omega}0.159$

From the curves of Figures 4 and 5 it is seen that the corresponding transfer function of the pulse former is approximately $1/\underline{-28}$ degrees. Thus the magnitude characteristic of the continuous portion of the system is relatively unaffected, while the necessary phase shift correction can be added as shown in Figure 7. The foregoing outlined procedure can be repeated for all frequencies of interest as is done in Figure 7.

The open loop characteristics of the sampling servo for various gains are transferred to the Nichols charts in Figure 8 to facilitate determination of closed loop performance. It is apparent that with unity amplifier gain ($K_v=0.5$) and continuous error detection and control, the response is overdamped. However, when this system is pulsed at 2 second intervals, the response becomes oscillatory.

The optimum value of amplifier gain can be selected in the conventional manner using Figure 8. Once a value of system gain is established, approximate values of system resonant frequency and maximum gain for sinusoidal inputs can be obtained easily from the curves of Figure 8. These values, in turn, yield a fairly accurate estimate of system damping and relative stability.

To verify the results obtained by the preceding method, a theoretical point-by-point transient solution employing a zero input and an initial output misalignment of unity was carried out for three values of amplifier gain. The results of this analysis are given in Figure 9. Inspection of Figures 8 and 9 will show that the approximation suggested in this paper can well be used to predict frequency of oscillation, relative overshoot, and solution time for the system.

Conclusions

It has been shown that quantitative analysis and design of sampling servos is possible, using frequency response techniques in conjunction with a uniquely defined transfer function for the usual switch-and-clamper combination. Limited application of the general approach outlined in the paper indicates that the performance characteristics of sampling or pulsed servos can be obtained with normal engineering accuracy. Inasmuch as the designer usually expects to obtain from the frequency response relative stability information and approximate values of resonant frequencies, rather than accurate solutions of the differential equations of the system, the use of this method over the range recommended is felt to be completely permissible and adequate.

Appendix I

The solutions for A_1 and B_1 are effected as follows

$$A_1=\frac{\theta_M}{\pi}\sum_{m=1}^{n}\sin\left[\phi+\frac{2\pi(m-1)}{n}\right]\times$$

$$\int_{\phi+\frac{2\pi(m-1)}{n}}^{\phi+\frac{2\pi m}{n}}\cos\epsilon d\epsilon$$

$$=\frac{\theta_M}{\pi}\sum_{m=1}^{n}\sin\left[\phi+\frac{2\pi(m-1)}{n}\right]\times$$

$$\left\{\sin\left(\phi+\frac{2\pi n}{n}\right)-\sin\left[\phi+\frac{2\pi(m-1)}{n}\right]\right\}$$

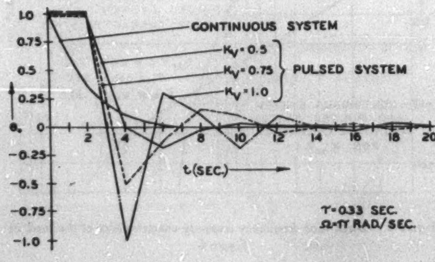
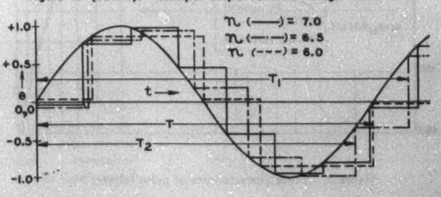


Figure 9 (left). Accurate theoretical response of the loop of Figure 6 to a zero input and an initial unity output

Figure 10 (below). Clamper output for a noninteger value of n



n	C1	C2	C3	C4	C5	C6	C7	C8	C9
10	0.984K	0	0	0	0	0	0	0	0.109K
9	0.980K	0	0	0	0	0	0	0.123K	0
8	0.976K	0	0	0	0	0	0.139K	0	0.108K
7	0.967K	0	0	0	0	0.161K	0	0.121K	0
6	0.955K	0	0	0	0.191K	0	0.136K	0	0
5	0.936K	0	0	0.234K	0	0.156K	0	0	0.104K
4	0.899K	0	0.300K	0	0.180K	0	0.129K	0	0.100K

WHERE K = θ_M

Figure 11. Harmonic distribution of clamper output

$$A_1=\frac{\theta_M}{\pi}\left\{\sum_{m=1}^{n}-\sin^2\left[\phi+\frac{2\pi(m-1)}{n}\right]+\right.$$

$$\sum_{m-1}^{n}\left[\left(\cos\frac{2\pi}{n}\right)\sin^2\left(\phi+\frac{2\pi m}{n}\right)-\right.$$

$$\left(\sin\frac{2\pi}{n}\right)\sin\left(\phi+\frac{2\pi m}{n}\right)\times$$

$$\left.\left.\cos\left(\phi+\frac{2\pi m}{n}\right)\right]\right\}$$

$$=\frac{\theta_M}{\pi}\left\{\left(\cos\frac{2\pi}{n}-1\right)\times\right.$$

$$\sum_{m=1}^{n}\left[\frac{1}{2}-\frac{1}{2}\cos 2\left(\phi+\frac{2\pi m}{n}\right)\right]-$$

$$\left.\sin\frac{2\pi}{n}\sum_{m=1}^{n}\frac{1}{2}\sin\left[2\left(\phi+\frac{2\pi m}{n}\right)\right]\right\}$$

A well-known theorem states that the sum of the sines (or cosines) of n angles given by $\theta_k=2\pi k/n$, where k takes on the values $1, 2, 3 \ldots, n$, is zero. Hence, for all integer values of n except 1 and 2, the expression for A_1 reduces to

$$A_1=\frac{n\theta_M}{\pi}\sin^2\left(\frac{\pi}{n}\right)$$

In a similar manner it can easily be shown that the reduced expression for B_1 is

$$B_1=\frac{n\theta_M}{2\pi}\sin\left(\frac{2\pi}{n}\right),\quad n=3, 4, 5, 6\ldots$$

Appendix II. Instances in Which n is a Noninteger

The expressions for A_1 and B_1 developed in Appendix I provide information about the operation of the pulsed system for any integer value of n except $n=1$ and $n=2$. However, the instances in which n is not an integer also should be investigated before attempting to specify the frequency response of the pulse former. In Figure 10, an output wave form for $n\cong 6.5$ ($n=360$ degrees/55 degrees) is shown in comparison with output wave forms for $n=6$ and $n=7$. Investigation of these wave forms reveals that for $k_i<n<k_j$ where k_i is any integer and $k_j=k_i+1$, the maximum deviation of the output wave form from that for the nearest value of k increases with a decrease in k_i. Furthermore, this deviation is a maximum for any given k_i when $2\pi/n-2\pi/k_j=1/2(2\pi/k_i-2\pi/k_j)$ or $k_ik_j/n-k_i=1/2$ since then the output wave form for n falls midway between the wave forms for k_i and k_j. (Here it is assumed that switching is commenced at the same instant in all three cases.) To determine the period T_1 of the output wave, consider the case where $\phi=0$. The value of T_1 is then defined as the time required for the output to assume one set of positive and negative values, ending at the sampling period nearest $t=T$ for which the magnitude of the pulse former output is a minimum. For example, in Figure 10 the period of the output wave form for $n\cong 6.5$ is that indicated by T_1 rather than T_2 since $|\theta_{PF(T1)}|<|\theta_{PF(T2)}|$. For each interval $k_i<n<k_j$, T_1 varies from T to some maximum value, the period of the output for $n=k_i$ and $n=k_j$ being T.

The relative deviation in period $(T-T_1)/T$ is a measure of the deviation of the output wave form from that upon which the calculations of A_1 and B_1 were based. For $n-k_1<k_j-n$, $(T-T_1)/T=(n-k_i)/n$; and for $k_j-n<n-k_i$, $(T-T_i)/T=(k_j-n)/n$. In general, then, for any value of n, $|T-T_1|/T=|n-k_c|/n$ where k_c is the integer closest to n in value. It is obvious that the relative period deviation for any given $|n-k_c|$ is inversely proportional to n. It is also apparent that $|n-k_c|$ can never exceed 0.5. Thus, as n approaches infinity, the transfer function for any value of n approaches that for nearby integer values of n. It may be shown easily that for a 10-per-cent deviation in period, $n=4.545$. This value of n has been used as a suggested upper frequency limit for application of this method.

Appendix III. Derivation of General Expressions for A_p and B_p

Before the study of the pulse former transfer function can be considered complete, the higher harmonics in the output wave form, which have been neglected in the body of this paper, must be investigated. A derivation of the general expressions for the sine and cosine components of these higher harmonics follows, as does a chart showing the frequency composition of the output for various integer values of n (Figure 11). It follows from this chart that, for integer values of n, the higher harmonics are small compared with the fundamental. Furthermore, in many systems one or more stages of low-pass filtering exist between the output of the pulse former and the controlled variable. Since the wave form at the output of the pulse former is usually of minor interest in itself, little harm is done in neglecting all but its fundamental component. This holds true even for cases in which the higher harmonics are of the same order of magnitude as the fundamental, provided sufficient filtering exists after the output of the pulse former.

For any integer value of n, the coefficient of the cosine component of the pth harmonic of the pulse former output wave form is given by

$$Ap=\frac{\theta_M}{\pi}\sum_{m=1}^{n}\sin\left[\phi+\frac{2\pi(m-1)}{n}\right]\times$$

$$\int_{\phi+\frac{2\pi(m-1)}{n}}^{\phi+\frac{2\pi m}{n}}\cos p e d e$$

$$Ap=\frac{\theta_M}{p\pi}\sum_{m-1}^{n}\sin\left[\phi+\frac{2\pi(m-1)}{n}\right]\times$$

$$\left\{\sin p\left[\phi+\frac{2\pi m}{n}\right]-\right.$$

$$\left.\sin p\left[\phi+\frac{2\pi(m-1)}{n}\right]\right\}$$

Through the use of the familiar trigonometric identities, this gives

$$Ap=\frac{\theta_M}{p\pi}\sum_{m=1}^{n}\left\{\sin\left[\phi+\frac{2\pi m}{n}\right]\cos\frac{2\pi}{n}-\right.$$

$$\left.\cos\left[\phi+\frac{2\pi m}{n}\right]\sin\frac{2\pi}{n}\right\}\times$$

$$\sin p\left[\phi+\frac{2\pi m}{n}\right]-\frac{\theta_M}{2p\pi}\times$$

$$\sum_{m=1}^{n}\left[\cos\left\{[1-p]\left[\phi+\frac{2\pi(m-1)}{n}\right]\right\}-\right.$$

$$\left.\cos\left\{[1+p]\left[\phi+\frac{2\pi(m-1)}{n}\right]\right\}\right]$$

or

$$Ap=\frac{\theta_M}{2p\pi}\left\{\left[\cos\frac{2\pi}{n}-1\right]\times\right.$$

$$\left.\sum_{m-1}^{n}\cos\left[(1-p)\left(\phi+\frac{2\pi m}{n}\right)\right]\right\}+$$

$$\frac{\theta_M}{2p\pi}\left\{\left[1-\cos\frac{2\pi}{n}\right]\times\right.$$

$$\left.\sum_{m=1}^{n}\cos\left[(1+p)\left(\phi+\frac{2\pi m}{n}\right)\right]\right\}-$$

$$\frac{\theta_M}{2p\pi}\sin\frac{2\pi}{n}\sum_{m=1}^{n}\sin\left[(1+p)\times\right.$$

$$\left.\left(\phi+\frac{2\pi m}{n}\right)\right]+\frac{\theta_M}{2p\pi}\sin\frac{2\pi}{n}\times$$

$$\sum_{m=1}^{n}\sin\left[(1-p)\left(\phi+\frac{2\pi m}{n}\right)\right]$$

In a similar manner it is simple to show that

$$Bp=\frac{\theta_M}{2p\pi}\left\{\left[1-\cos\frac{2\pi}{n}\right]\times\right.$$

$$\left.\sum_{m=1}^{n}\sin\left[(1+p)\left(\phi+\frac{2\pi m}{n}\right)\right]\right\}+$$

$$\frac{\theta_M}{2p\pi}\left\{\left[1-\cos\frac{2\pi}{n}\right]\times\right.$$

$$\left.\sum_{m=1}^{n}\sin\left[(1-p)\left(\phi+\frac{2\pi m}{n}\right)\right]\right\}+$$

$$\frac{\theta_M}{2p\pi}\sin\frac{2\pi}{n}\sum_{m=1}^{n}\cos\left[(1+p)\times\right.$$

$$\left.\left(\phi+\frac{2\pi m}{n}\right)\right]+\frac{\theta_M}{2p\pi}\sin\frac{2\pi}{n}\times$$

$$\sum_{m=1}^{n}\cos\left[(1-p)\left(\phi+\frac{2\pi m}{n}\right)\right]$$

Although these expressions appear to be lengthy and involved, computation reveals that the magnitudes of the higher harmonics also are independent of the phase of the input wave.

In the chart of Figure 11 are shown the relative magnitudes of the first nine harmonics as computed from the expressions just derived. In this figure, $Cp=\sqrt{Ap^2+Bp^2}$ is the magnitude of the pth harmonic.

References

1. Theory of Servomechanisms, **H. L. Hazen.** *Journal*, Franklin Institute (Philadelphia, Pa.), volume 218, 1934, pages 279–331.

2. Theory of Servomechanisms (book), **Hubert M. James, Nathaniel B. Nichols, Ralph S. Phillips.** McGraw-Hill Publishing Company, Inc., New York, N. Y., chapter 5, 1947.

3. A Servo System Operating on Discontinuous Information, **C. Holt Smith, D. Lawden.** Military College of Science (England), 1950.

4. Sampled-Data Control Systems Studied Through Comparison of Sampling with Amplitude Modulation, **William K. Linvill.** *AIEE Transactions*, volume 70, part II, 1951, pages 1779–86.

Discussion

John R. Ragazzini and **Lotfi A. Zadeh** (Columbia University, New York, N. Y.): The approximation described by the authors is of limited applicability since it is valid only where the sampling frequency is high enough to permit the replacement of a sampled-data device by a continuous one. In the opinion of the discussers, a shortcoming of this paper is that no general criteria of the goodness of the approximation are provided. In fact, only comparisons with exact results can assure the designer that the approximation in question is applicable to the problem at hand.

It is observed that the authors' results can be obtained more simply if use is made of well-known techniques of analysis and synthesis of sampled-data systems.[1,2] Referring to Figure 1 of the discussion, an approximate transfer function, θ_C/θ_A, is obtained, defined as the ratio of the fundamental component of the output to a unit sinusoidal input. The input is expressed in exponential form as

$$\theta_A(t)=A\epsilon^{j\omega t} \tag{1}$$

Linvill[1] has shown that when this function is sampled by means of the switch S, the sampled function $\theta_A{}^*$ is expressible as an infinite series where

$$\theta_A{}^*(t)=\frac{A}{T}\sum_{n=-\infty}^{+\infty}\epsilon^{j(\omega+n\omega_s)t} \tag{2}$$

ω_s is the sampling frequency. The discussers[2] have shown that the transfer function of a clamp circuit is

$$F_I(s)=\frac{1}{s}(1-\epsilon^{-sT}) \tag{3}$$

where T is the sampling interval. From equation 3, the frequency response function of the clamp circuit is

$$F_I(j\omega)=\frac{1}{j\omega}(1-\epsilon^{-j\omega T}) \tag{4}$$

Taking the exponential series given in equation 2 and operating on it with the transfer function of the clamp circuit, there results at the output

$$\theta_B(t)=\frac{A}{T}\sum_{n=-\infty}^{+\infty}\frac{1}{j(\omega+n\omega_s)}\times$$

$$(1-\epsilon^{-j(\omega+n\omega_s)T})\epsilon^{j(\omega+n\omega_s)t} \tag{5}$$

If the fundamental term only is retained in this expression, the transfer function as defined by the authors becomes

$$G(j\omega)=\frac{1}{j\omega T}(1-\epsilon^{-j\omega T}) \tag{6}$$

This simple expression is equivalent to the result obtained by the authors.

The foregoing derivation is given mainly to show that there are simpler methods of obtaining a transfer function in the sense used by the authors. In addition, the discussers would like to clarify the impression given by the authors that exact methods are overly tedious and lengthy. Due to its greater complexity, a sampled-data system requires a somewhat more lengthy analysis than its continuous counterpart. However, properly employed, neither the frequency analysis method nor the z-transform method is significantly more difficult to use than the approximation used by the authors. In fact, if approximations are to be employed, they can be obtained more readily directly from the exact methods.

References

1. Sampled-Data Control Systems Studied Through Comparison of Sampling with Amplitude Modulation, **W. K. Linvill.** *AIEE Transactions*, volume 70, part II, 1951, pages 1779–88.

2. The Analysis of Sampled-Data Systems, **J. R. Ragazzini, L. A. Zadeh.** *AIEE Transactions*, volume 71, part II, November 1952, pages 225–34.

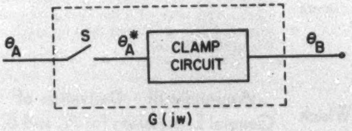


Figure 1. Schematic diagram of sampled-data transmission element

Improving the Dynamic Response of Airplanes by Means of Electric Equipment

WILLIAM R. MONROE
NONMEMBER AIEE

Synopsis: This paper begins with an elementary discussion of the mechanics of airplane flight and control. Following this discussion electrical methods for improving the transient response of an airplane in the yawing and pitching modes are outlined and illustrated by two examples. The paper is concluded by a brief discussion of a method by which an airplane cable control system can be replaced by an electric servomechanism. Results of analyses using differential equations and Nyquist plots are discussed and compared with results obtained on an analogue computer.

SUBSEQUENT to the first World War the airplane has undergone many remarkable changes. For example, during World War I the operating range of airplanes was up to an altitude of 12,000 feet and a speed of 188 miles per hour. By the end of World War II the operating altitude and speed of airplanes were increased to 25,000 feet and 450 miles per hour.

In the Korean war the operating altitude has been extended to 50,000 feet and the speed of 670 miles per hour. This increasing performance has introduced a new set of airplane control problems, some of which are being solved by the use of electric equipment which supplements the aerodynamic characteristics of the airplane. Some of the design problems that would be encountered if a World War II airplane were to fly at increasingly high speeds are:

1. Extreme variations in control surface sensitivity.
2. Peculiar force feel characteristics that might even require the pilot to move the control column in an illogical direction.
3. Wide variation in airplane damping.
4. Extremely large pilot forces required to control the airplane.

The solution of these problems can be obtained by wider use of electric equipment in the airplane control system. Ways in which this is being accomplished are discussed in this paper.

Nomenclature

A = wing area, square feet
α = Angle of attack, degrees = angle between relative wind and wing chord
β = aerodynamic yaw, degrees = angle between the plane of symmetry and the relative wind projected into the XY plane
θ = pitch angle, degrees = angle between the F.R.L. and horizon
ϕ = roll angle, degrees = angle between the horizon and the Y axis
ψ = angle of geometric yaw, degrees = angle between F.R.L. and a reference heading
γ = climb angle, degrees = angle between relative wind and the horizon
F.R.L. = fuselage reference line
X axis = the axis, in the plane of symmetry, about which an airplane rolls
Y axis = the axis perpendicular to the plane of symmetry about which an airplane pitches
Z axis = the axis, in the plane of symmetry perpendicular to the Y axis, about which an airplane yaws
g = gravity constant = 32 feet per second per second
s = Laplace transform variable
L = lift force, pounds
F_P = pilot force, pounds
F_b = force on control system due to bobweight, pounds
n_z = normal acceleration, g's
W = weight of airplane, pounds
X_t = servomechanism input deflection, inches
ω_n = natural frequency, radians per second
ζ = damping ratio
δ_a = angle of aileron deflection, degrees
δ_r = angle of rudder deflection, degrees
δ_s = angle of stabilizer deflection, degrees
μ_T = total open loop-transfer function of stick free longitudinal control system
m, q, c, b, V, C_{L_q}, C_{L_α}, $C_{L_{d\alpha/dt}}$, $C_{z_{\delta_s}}$, C_{m_α}, $C_{m_{d\alpha/dt}}$, $C_{m_{\delta_s}}$, C_{m_q}, I_x, I_y, I_z, I_{xz}, C_{l_p}, C_{l_r}, C_{l_β}, $C_{l_{\delta_a}}$, $C_{l_{\delta_r}}$, C_{n_p}, C_{n_r}, C_{n_β}, $C_{n_{\delta_a}}$, $C_{n_{\delta_r}}$, C_{y_p}, C_{y_r}, C_{y_β}, $C_{y_{\delta_r}}$ and C_{L_0} are defined in references 1, 2, and 3.

How an Airplane Is Controlled

If an air stream is directed on an airfoil or an airplane wing, a lifting force as shown in Figure 1 is created. This lift force depends on the shape and size of the airfoil, the air speed, the air density (altitude), and the angle of attack, that is, the angle between the direction of the air stream and the chord of the airfoil. For a given airplane, speed, and altitude, the lift force varies with angle of attack as shown in Figure 1. If the lift force exceeds the weight of the airplane, the airplane climbs, but if the lift is less than the weight, the airplane descends.

In flight, if the pilot wants to climb, he does so by increasing the lift force which is accomplished by increasing the angle of attack of the airplane wing. The manner in which this is done is shown in Figure 2. For an airplane in level flight, the lift on the wing airfoil and tail airfoil are balanced and the lift is just equal to the airplane weight. When the pilot exerts a backward force F_p on the control stick, he rotates the stabilizer through an angle $-\delta_s$ as shown in Figure 2. This decreases the stabilizer angle of attack α_T which reduces the tail lift and therefore applies a nose-up moment M to the airplane. By applying Newton's Law to the airplane, the relation between the inertia and aerodynamic forces and moments can be determined. From the resulting equations, the airplane angular rotation θ with respect to the horizon and the wing angle of attack α_w resulting from the surface deflection, δ_s can be calculated. The normal acceleration n_z and the flight path angle γ can then be computed.[1,2]

For example, if the airplane is assumed flying level at constant thrust and if the speed changes (ineffective for periods of the order of magnitude of 1 second or less) are neglected, the airplane motion can be expressed by the lift force and moment equations 1 and 2 given in the following text.

Lift force equation:

$$\left[\left(-\frac{mV}{qA}+\frac{c}{2V}C_{L_q}\right)s\right]\theta+\left[\left(\frac{mV}{qA}+\frac{c}{2V}C_{L_{d\alpha/dt}}\right)s+C_{L_\alpha}\right]\alpha=C_{Z_{\delta_s}}\delta_s \tag{1}$$

Moment equation:

$$\left[\frac{I_y}{qAc}s^2-\left(\frac{c}{2V}C_{m_q}\right)s\right]\theta+\left[\left(-\frac{c}{2V}C_{m_{d\alpha/dt}}\right)s-C_{m_\alpha}\right]\alpha=C_{m_{\delta_s}}\delta_s \tag{2}$$

In these equations for a given airplane flying at a given speed and altitude, all of

---

Paper 52-338, recommended by the AIEE Air Transportation Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Middle Eastern District Meeting, Toledo, Ohio, October 28-30, 1952. Manuscript submitted July 30, 1952; made available for printing September 23, 1952.

William R. Monroe is with North American Aviation, Inc., Los Angeles, Calif.

The author wishes to acknowledge the co-operation rendered by his associates during the preparation of this paper. In particular the work of R. G. Bang, W. Lobitz, and H. D. Depew in setting up the experimental equipment, conducting tests, and preparing the illustrations is greatly appreciated.

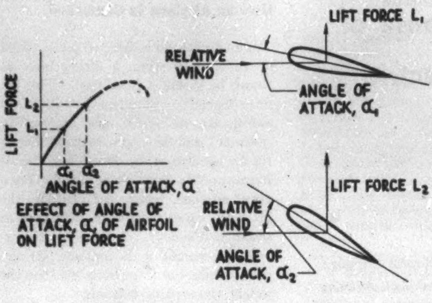


Figure 1. Lift force on an airplane airfoil

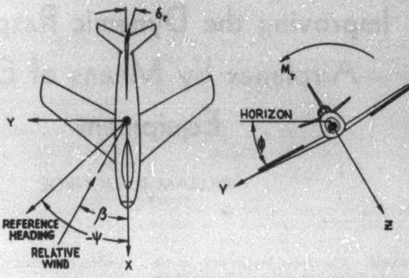


Figure 3. Airplane lateral—directional control parameters

the symbols may be regarded as constants with the exception of θ, α, δ_s, and the Laplace operator[3] s. The equations may be solved for the airplane pitch angle θ or angle of attack α response to control surface input δ_s. The response in normal acceleration n_z can be obtained by converting pitch angle and angle of attack to normal acceleration by the equation

$$n_z g = -V\left(\frac{d\theta}{dt}-\frac{d\alpha}{dt}\right)-g \qquad (3)$$

The responses of the airplane in roll and yaw can be obtained in a similar manner. The lateral-directional co-ordinate system is defined in Figure 3 and the equations of motion[4] are given here.

Rolling moment equation:

$$\left[\frac{I_x}{qAb}s^2-\frac{b}{2V}C_{l_p}s\right]\phi+\left[-\frac{I_{xz}}{qAb}s^2-\frac{b}{2V}C_{l_r}s\right]\psi-[C_{l_\beta}]\beta=C_{l_{\delta_a}}\delta_a+C_{l_{\delta_r}}\delta_r \qquad (4)$$

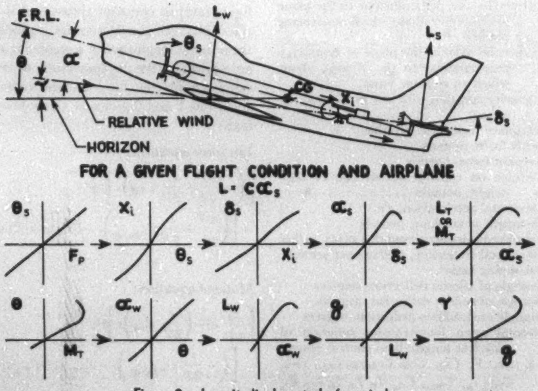


Figure 2. Longitudinal control of an airplane

These curves are presented as visual aids to discussions in the text and are not intended to represent the exact functions which may exhibit infinitely many variations in shape

Yawing moment equation:

$$\left[-\frac{I_{xz}}{qAb}s^2-\frac{b}{2V}C_{n_p}s\right]\phi+\left[\frac{I_z}{qAb}s^2-\frac{b}{2V}C_{n_r}s\right]\psi-[C_{n_\beta}]\beta=C_{n_{\delta_a}}\delta_a+C_{n_{\delta_r}}\delta_r \qquad (5)$$

Lateral force equation:

$$\left[-\frac{b}{2V}C_{y_p}s-C_{L_0}\right]\phi+\left[\frac{mV}{qA}s-\frac{b}{2V}C_{y_r}s\right]\psi+\left[\frac{mV}{qA}s-C_{y_\beta}\right]\beta=C_{y_{\delta_r}}\delta_r \qquad (6)$$

Improving the Oscillatory Yawing Motion of an Airplane

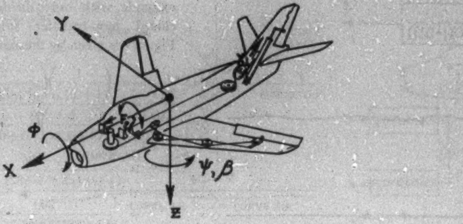
If the airplane dynamic response to a pilot force input is not satisfactory, the response can be improved by adding electrical components to the control loop. As a first example, consider a lateral control problem of the airplane shown in Figure 4.

The airplane is controlled directionally by the pilot deflecting the rudder pedals. As shown in the block diagram of Figure 4, when the pilot exerts a force F_P on the rudder pedals, he causes an input X_i to the rudder servo mechanism which in turn deflects the rudder through an angle δ_r. This rudder deflection causes the airplane to yaw and the pilot alters his pedal force until the yaw angle has the desired value.

Inputs other than those of the pilot can occur as shown by the rolling gust in Figure 4. Because of aerodynamic cross coupling a rolling gust can cause the airplane to yaw as well as roll.

To reduce the difficulty of flying a modern, complex airplane, it is desirable that the pilot be freed from the necessity of stabilizing an underdamped yawing oscillation, such as that which could be excited by a rolling gust.

There are two obvious ways to change the dynamics of an airplane flight control system. In one method the effective surface deflection δ_r can be changed not only by pilot inputs but by feedback signals proportional to aerodynamic variables. With the other method, the forces experienced by the pilot can be made proportional to various aerodynamic parameters. Five ways of accomplishing these two types of changes are shown in Figure 5. The middle sketch shows schematically how a force proportional to airplane pitch acceleration could be added to the control column. The other four show ways in which the effective surface deflection could be made proportional to various feedback signals as well as pilot inputs. These feedback signals could be proportional to any combination of aerodynamic or mechanical parameters selected to improve the transient response of the flight control system. In the first example discussed in this paper, the



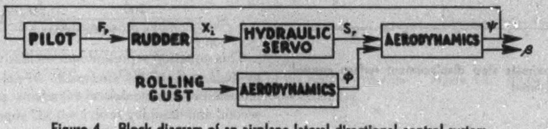

PILOT
F_r
RUDDER
X_i
HYDRAULIC SERVO
S_r
AERODYNAMICS
ψ
β
ROLLING GUST
AERODYNAMICS
φ

Figure 4. Block diagram of an airplane lateral directional control system

methods shown on the right side of Figure 5 will be considered. Several possible aerodynamic parameters which can be fed back to improve the transient response are shown in Figure 6.

In this figure the aerodynamic transfer function β/δ_r has been simplified for illustrative reasons to the second order system expressed by the equation

$$\frac{\beta(s)}{\delta_r(s)}=\frac{20}{s^2+0.87s+20} \tag{7}$$

The transfer function of the rudder servomechanism shown in Figure 6 is assumed to be given by the relation

$$\frac{\delta_r(s)}{X_i(s)}=\frac{1}{1+0.1s} \tag{8}$$

By alternately setting the constants C_1, C_2, and C_3 equal to zero the effects of β, $d\beta/dt$, and $d^2\beta/dt^2$ feedback can be investigated separately.

If all three of these constants are zero, there is no feedback and the sideslip response β of the airplane to a pilot step input at X_o is described by the relation

$$\beta(s)=\frac{20C_o}{(1+0.1s)(s^2+0.87s+20)s} \tag{9}$$

This describes an oscillation composed of a nonoscillatory damped mode with a time constant $\tau=0.1$ second on which is superimposed an oscillatory mode with a natural frequency of $\omega_n=\sqrt{20}=4.47$ radians per second and a damping ratio given by the relation $2\zeta\omega_n=0.87$ or, that is $\zeta=0.097$. This analytical result is verified by the transient response (Figure 7) of this system obtained from an analogue computer.

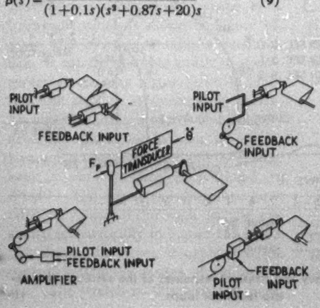
If some means of detecting and feeding back sideslip β is provided, the surface position will be proportional to pilot input and β. This feedback signal proportional to β then could be differentiated electrically and any derivative used to move the surface in a proportional manner. Such an arrangement is shown in Figure 8. Here the sideslip angle β is computed by means of an aerodynamic equation relating β with geometric yaw angle ψ, lateral acceleration n_β, and the airplane weight W, that is

$$\beta=n_\beta\left\{q_cA\left(C_{y_\beta}-\frac{b}{l_R}C_{n_\beta}\right)\right\}^{-1}W+\frac{1}{q_cA}\left[\frac{I_y/l_R}{C_{y_\beta}-\frac{b}{l_R}C_{n_\beta}}\right]\frac{d^2\psi}{dt^2} \tag{10}$$

The geometric yaw rate $d\psi/dt$ can be detected by a rate gyro from which the signal can be differentiated to obtain $d^2\psi/d^2t$. This signal is multiplied electrically by the quantity $1/q_cA[(I_y/l_R)/(C_{y\beta}-b/l_RC_{n\beta})]$. As shown in Figure 8 the signal is added to another signal which is proportional to the product of the computed airplane weight and the lateral acceleration measured by an accelerometer. The resulting value of β can be fed back alone or differentiated electrically as shown in the figure.

The effect of these three feedback signals can be investigated by combining the transfer functions of the forward loop β/X_i and the feedback loops C_1X_i/β, $C_2X_i/(d\beta/dt)$, and $C_3X_i/(d^2\beta/dt^2)$, all of which are shown plotted in Figure 9. Combining the transfer functions of the forward and feedback loops provides the transfer function and transient responses of Figures 10 and 11. In Figure 10 the time constant of the instrumentation and feedback paths are assumed to be very small, that is, $\tau=0.01$ second, but in Figure 11 this time constant is assumed to be 0.11 second.

Referring to Figure 10 it is clear that with β feedback of gain larger than 0.8 the Nyquist plot encircles the $-1+j0$ point and hence[5] the system would be unstable. The diverging oscillation is shown in the transient response of Figure 10. The stabilizing effect of $d\beta/dt$ feedback can be seen by reference to curves 2A and 2B from which it is clear that $d\beta/dt$ feedback is definitely stabilizing. Although the Nyquist plot with $d^2\beta/dt^2$ feedback lies to the right of


PILOT INPUT
FEEDBACK INPUT
FORCE TRANSDUCER
$\ddot{\delta}$
F_p
PILOT INPUT
FEEDBACK INPUT
PILOT INPUT
FEEDBACK INPUT
AMPLIFIER
FEEDBACK INPUT
PILOT INPUT

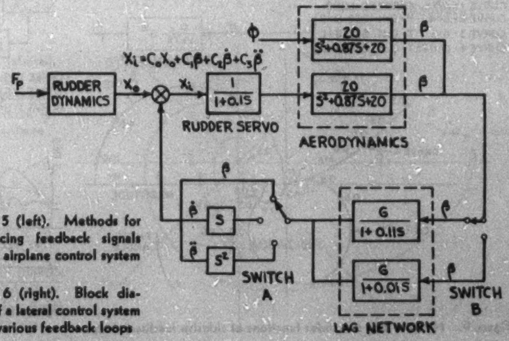

φ
$X_i=C_oX_o+C_1\beta+C_2\dot{\beta}+C_3\ddot{\beta}$
F_p
RUDDER DYNAMICS
X_o
X_i
$\frac{1}{1+0.1S}$
RUDDER SERVO
$\frac{20}{S^2+0.87S+20}$
$\frac{20}{S^2+0.87S+20}$
β
β
AERODYNAMICS
β
$\dot{\beta}$
S
$\ddot{\beta}$
S^2
SWITCH A
$\frac{G}{1+0.11S}$
$\frac{G}{1+0.01S}$
β
β
LAG NETWORK
SWITCH B

Figure 5 (left). Methods for introducing feedback signals into an airplane control system

Figure 6 (right). Block diagram of a lateral control system with various feedback loops

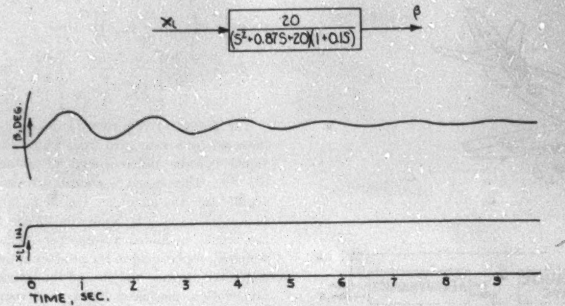
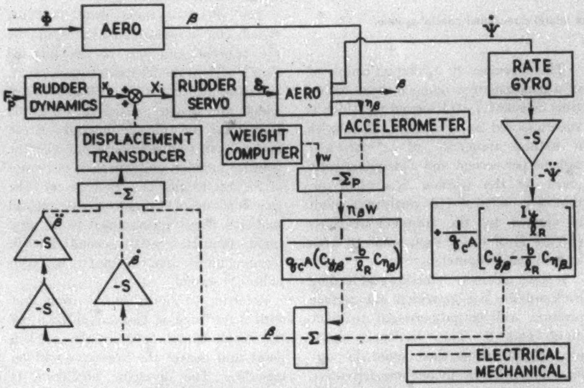


Figure 7. Airplane sideslip response to an approximate step displacement surface control servomechanism input



Φ
AERO
β
$\dot{\Psi}$
RUDDER DYNAMICS
F_P
x_b
x_i
RUDDER SERVO
δ_r
AERO
β
n_β
RATE GYRO
DISPLACEMENT TRANSDUCER
WEIGHT COMPUTER
W
ACCELEROMETER
-S
$-\ddot{\Psi}$
-Σ
$-\Sigma_P$
$n_\beta W$
$\ddot{\beta}$
$\dot{\beta}$
-S
-S
-S
$\frac{1}{q_c A\left(C_{y_\beta}-\frac{b}{\ell_R}C_{n_\beta}\right)}$
$+\frac{1}{q_c A}\left[\frac{\frac{I_\psi}{\ell_R}}{C_{y_\beta}-\frac{b}{\ell_R}C_{n_\beta}}\right]$
β
-Σ
ELECTRICAL
MECHANICAL

Figure 8. Mechanization of sideslip feedback loops

the $d\beta/dt$ feedback plot, the transient response of the $d^2\beta/dt^2$ feedback system is less damped than the system with $d\beta/dt$ feedback.

The added time lag in instrumentation and feedback paths is destabilizing as indicated in Figure 11.

The results obtained by analogue computer also can be obtained analytically with this simplified system. For example with only $d\beta/dt$ feedback the closed loop-transfer function β/x_0 of Figure 6 is given by the expression

$$\frac{\beta(s)}{X_0(s)}=\frac{\left(\frac{1}{1+0.1s}\right)\left(\frac{20}{s^2+0.87s+20}\right)}{1+\left(\frac{1}{1+0.1s}\right)\left(\frac{20}{s^2+0.87s+20}\right)(s)}$$

$$=\frac{20}{(1+0.1s)(s^2+0.87s+20)+20s}$$

$$=\frac{200}{s^3+10.87s^2+28.7s+200}$$

$$=\frac{200}{(s+0.91)(s^2+9.96s+220)} \qquad (11)$$

This equation represents an oscillatory mode $[\beta_2=C_2\epsilon^{-(9.96/2)t}\sin(\sqrt{195}\ t+\phi)]$ of natural frequency $\omega_0=14.0$ radians per second and damping ratio $\zeta=0.337$ superimposed on a nonoscillatory mode $\beta_1=c_1\epsilon^{-0.91t}$ with positive damping and a time constant of $\tau=1/0.91=1.10$ seconds. The output resulting from a step input approaches a steady-state value of $\beta=\beta_0$. Thus the solution of the equation of motion resulting from a step input is given by the relation

$$\beta=c_1\epsilon^{-0.91t}+c_2\epsilon^{-4.98t}\sin(14.8t+\phi)+c_3 \qquad (12)$$

Comparison of this analytical result with its corresponding transient response (curve 28 of Figure 10) indicates that the computed time constant of 1.1 seconds occurs, but the underdamped oscillation is not apparent in the figure. This apparent discrepancy is due to the fact that the magnitude of the constants c_1 and c_2 in the preceding equations have not been discussed. These constants, determined by the boundary conditions $\beta=d\beta/dt=d^2\beta/dt^2=0$ at $t=0$ and $\beta=\beta_0$ at $t=\infty$, are $c_1=1.03$, $c_2=0.066$, and $c_3=1.00$.

Thus it is apparent that the magnitude of the oscillatory term is so small that it is not noticeable in comparison with the

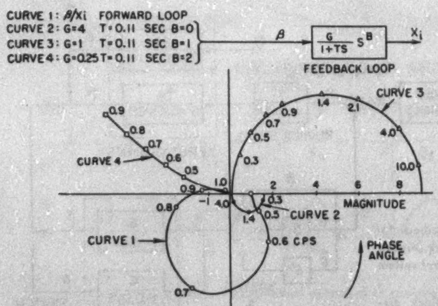


Figure 9. Nyquist plot of transfer functions of sideslip feedback loops

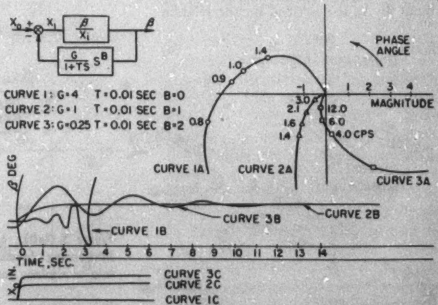


Figure 10. Frequency and transient responses of the system with sideslip feedback loops

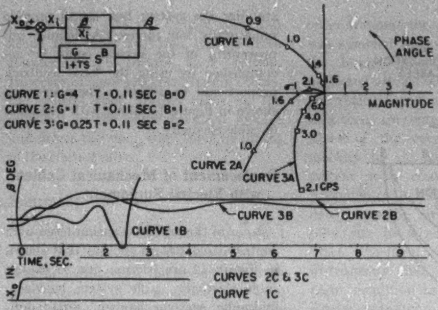
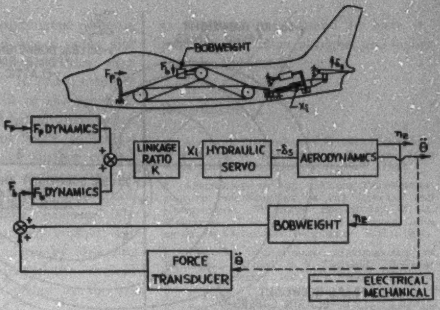


Figure 11. Frequency and transient responses of the system with sideslip feedback loops and additional feedback time lag



BOBWEIGHT
F_p
F_b
δ_s
X_i
F_p DYNAMICS
F_b DYNAMICS
LINKAGE RATIO K
X_i
HYDRAULIC SERVO
$-\delta_s$
AERODYNAMICS
n_z
$\ddot{\theta}$
BOBWEIGHT
n_z
FORCE TRANSDUCER
$\ddot{\theta}$
ELECTRICAL
MECHANICAL

Figure 12. Block diagram of the longitudinal flight contact system of an airplane

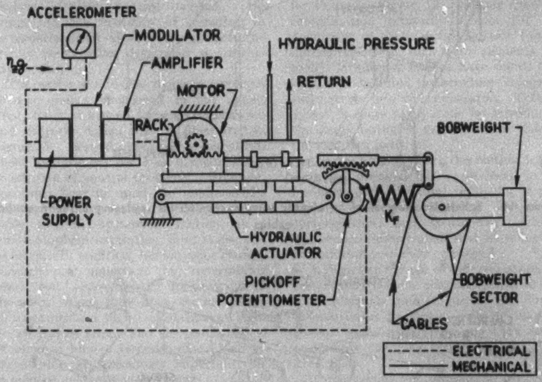
nonoscillatory mode.

From the analytical and analogue computer results it is evident that of the systems discussed the optimum system from the standpoint of dynamic overshoot is that with $d\beta/dt$ feedback.

Improving the Dynamic Response of a Longitudinal Flight Control System

The longitudinal control system shown in Figure 12 is operated when the pilot applies a force F_p to the control column of the airplane. This force and the impedance of the mechanical system connecting the control column with the hydraulic servomechanism determine the input deflection X_i to the servomechanism. This deflection and the static and dynamic characteristics of the hydraulic servomechanism determine the airplane control surface angle $-\delta_s$ which by means of the airplane inertia and aerodynamics determines the airplane normal acceleration n_z. To provide a system that feels natural, the pilot is provided with a control column force proportional to normal acceleration. This is accomplished by means of a bobweight which is shown in Figure 12. Since this system contains a feedback loop, it can be unstable and it is always necessary to make certain prior to flight that the airplane is dynamically stable. An unstable airplane can be stabilized by detecting pitch acceleration $d^2\theta/dt^2$ by means of an accelerometer and by feeding back a control force proportional to $d^3\theta/dt^3$. The force transducer can be a force servomechanism such as the one shown schematically in Figure 13. With this device, the force is applied to the control system by means of the spring K_F. The stretch in this spring is measured by the potentiometer and the resulting signal is compared with the signal, n_zg, from the accelerometer. If these signals are not equal, the motor drives the hydraulic servomechanism stretching the spring until the signal from the potentiometer is equal to the accelerometer signal.

As an example of the application of this method of improving the dynamic response of an airplane by means of electric equipment, consider the system represented by the Nyquist plot of Figure 14. With the original bobweight set at 35 pounds per g the system was dynamically unstable. A smaller bobweight could



ACCELEROMETER
n_zg
MODULATOR
AMPLIFIER
HYDRAULIC PRESSURE
RETURN
MOTOR
RACK
BOBWEIGHT
POWER SUPPLY
K_F
HYDRAULIC ACTUATOR
PICKOFF POTENTIOMETER
BOBWEIGHT SECTOR
CABLES
ELECTRICAL
MECHANICAL

Figure 13 (above). Mechanization of a transducer providing an output force for an input voltage

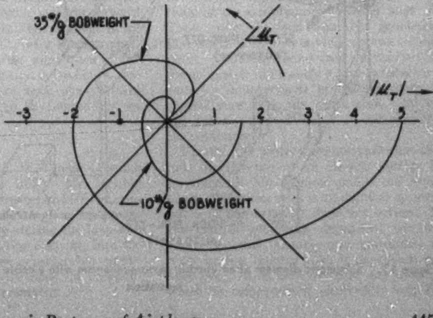


Figure 14 (right). Nyquist plot of a longitudinal flight control system

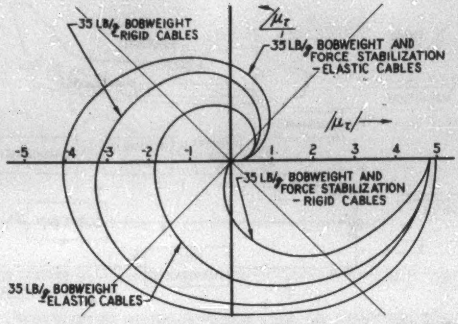


Figure 15. Nyquist plot of a longitudinal flight control with force feedback stabilization

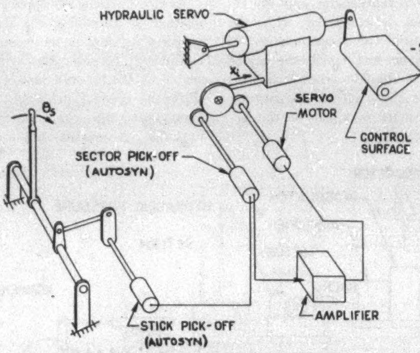


Figure 16. Schematic diagram of an electric servomechanism for replacing airplane control cables

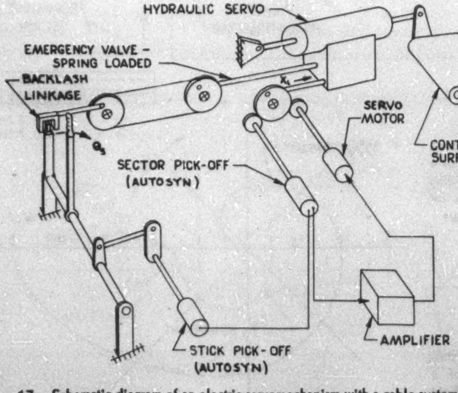


Figure 17. Schematic diagram of an electric servomechanism with a cable system for emergency operation

stabilize the system but would provide unnaturally low forces. A definite improvement of the system could be accomplished by using the force feedback proportional to $d^2\theta/dt^2$. The results are shown in Figure 15, along with the effect of cable elasticity.

Replacement of Mechanical Cables with Electric Equipment

Some of the problems encountered with control cable system such as that shown in Figure 12 are friction due to cables, cable elasticity, cable system backlash, hydraulic servomechanism, drift, difficulty in reducing control surface resolution, and difficulty of introducing feedback signals for autopilot operation and artificial stability devices.

These problems can be reduced considerably by the arrangement shown in Figure 16. With this system the control stick rotation θ_s is compared with the servomechanism input position X_i by means of two autosyns. The signal representing this difference is amplified and the servomechanism motor is operated until the servomechanism input X_i corresponds to the control stick deflection. With this system, the small friction of the stick pick-off autosyn is substituted for the relatively large friction of the cable system. The backlash between the stick and servomechanism input, governed by the hydraulic servomechanism input impedance, servomechanism motor output torque, and accuracy of the autosyns, can be reduced to a negligible value. Electric feedback signals from the autopilot or electric stability devices can be introduced easily into the amplifier and there is no necessity for extra servomechanism motors for these devices. By introducing an additional electric feedback from the control surface itself, surface resolution and drift can be improved.

One of the major problems with a system such as this electric servomechanism is fail safety. This problem can be solved by use of two duplicate independent electric systems or by an arrangement such as the one shown in Figure 17.

The system illustrated is the same as that of Figure 16 except that a cable system, emergency hydraulic valve, and cable backlash linkage have been added. This cable system does not add any friction during normal operations because the electric servomechanism motor drives the hydraulic servomechanism which drags the cables with it as it moves. This action keeps the control surface synchronized with the control stick and keeps the cable backlash linkage from contacting the control stick system. However, if some electrical failure occurs, the control stick system takes up the backlash in the cable backlash linkage and drives the emergency valve by means of the control cables.

Summary

In the foregoing discussion, the control of an airplane has been described and several methods of improving the transient response of the airplane have been mentioned. In the interest of clarity, the various components of the flight control system were simplified considerably. For example, the hydraulic servomechanisms were treated as simple first-order systems, the aerodynamic response was assumed to be of second order, and the human pilot characteristics were neglected altogether. In practice this is not realistic and most of the problems that occur are centered about these items. Most of the current work utilizes a mockup of the actual surface control system with control surface position presented to an analogue computer which determines the airplane response. The pilot sits in a simulated cockpit and "flies" the flight simulator by observing instruments that indicate the airplane flight condition as directed by the analogue computer.

The application of electric equipment in changing the dynamic response of airplane flight control systems is in its infancy and can be expected to grow rapidly as airplanes are required to have increased performance at higher and higher altitudes and speeds.

References

1. Airplane Performance Stability and Control (book), C. D. Perkins, R. E. Hage. John Wiley and Sons, Inc., New York, N. Y., 1950.

2. Application of the LaPlace Transformation to the Solution of the Lateral and Longitudinal Stability Equations, G. A. Mokrzycki. *TN 2002*, National Advisory Committee for Aeronautics (Washington, D.C.), 1950.

3. Modern Operational Mathematics in Engineering (book), R. V. Churchill. McGraw-Hill Book Company, Inc., New York, N. Y., 1944.

4. A Compilation of the Papers Presented at the NACA University Conference on Aerodynamics, June 21–23, 1948. Durand Reprinting Committee, California Institute of Technology (Pasadena, Calif.), 1948.

5. Servomechanisms and Regulating System Design (book), H. Chestnut, R. W. Mayer. John Wiley and Sons, Inc., New York, 1951.

Discussion

Max E. Gellert (Boeing Airplane Company, Seattle, Wash.): Damping an airplane by the addition of stabilizing signals is not a new concept. However, the method, shown here, of using sideslip (or its first or second derivative) is much more difficult than the author of this paper indicates.

In equation 10, an expression for computing β is presented, but no mention is made of the fact that q_c, C_{n_β}, and C_{y_β} are all dependent on air speed and altitude. Two quantities that the author considers constant, $-q_cA\,[C_{y_\beta}-b/l_RC_{n_\beta}]$ and $\dfrac{1}{q_cA}\left[\dfrac{I_y/l_R}{C_{y_\beta}-b/l_RC_{n_\beta}}\right]$, are really very complicated variables. Therefore a β computer is an expensive and difficult instrument to build. Also, variable n_β seems to be missing from equation 10, although it is shown in Figure 8 and referred to in the text.

Representation of an airplane by a second order differential equation is an oversimplification. Furthermore, the β response to a rudder deflection will not be the same as the β response to a roll angle as indicated in Figure 6, except by an almost impossible coincidence.

Referring to Figure 2, the author shows a plot of θ as a function of M_T. θ is always referred to the horizon so that it would be more correct to plot α_w directly as a function of M_T. It is possible to have $M_T=0$ with almost any value of θ. Similarly, γ is not a single valued function of g.

Except for oversights in Figure 9, where the curves are not labelled, and Figure 10, where curves 2B and 3B are apparently reversed, the servomechanism theory is straightforward and conventional.

S. Pines (Republic Aviation Corporation, Farmingdale, N. Y.): The use of analogue simulation of flight control problems has been an active practice in aircraft design and testing for the past 2 years. Quite a bit of the initial spade work in making this procedure a practical tool has been carried out by Mr. Monroe and his co-workers. The fruits of this effort have paid off in financial economy, in time saved in obtaining fixes, and perhaps even in lives.

In this paper, Mr. Monroe has clearly described the role of fundamental electrical engineering techniques in the synthesis of modern airplane control systems. He brings out the necessity of resorting to servomechanisms theory to extract the greater performance demanded of present-day aircraft.

The writer would like to place emphasis on two points mentioned in the paper. First, there is a need for good mechanism and gadgetry design such as illustrated in Figure 5. Second, the use of simplified linear analysis can serve as a useful guide to solution of control problems. However, one must bear in mind the nonlinearities inherent in the system such as control cable stretch, hysteresis, friction, dead zones, amplifier saturation, limit stops, and so forth. In addition, human pilot characteristics are important for determining what "feel" characteristics are desirable to the pilot. Along these lines, the utilization of mockups and flight simulation as mentioned in this paper, and the psychological problem of human response characteristics are playing an exceedingly important role in modern control system design.

In mentioning these items it becomes increasingly evident that a straight-forward analytical approach to the problem is almost beyond the ability of a modern engineering staff. The great advantage of the simulator technique is that it enables a convenient combination of both physical and analytical analogues to be compatibly set up and used.

M. U. Clauser and **J. G. Truxal** (Purdue University, Lafayette, Ind.): Mr. Monroe has brought to the aerodynamicist the concept of improving the airplane controllability and stability by closing the loop and introducing compensating elements into the loop, and he deserves considerable credit. The means by which he does the compensating deserves some comment.

The closed loop but uncompensated system and its transfer function can be viewed as having certain undesirable characteristics, and these characteristics are to be eliminated (or modified) by the use of compensating elements. Mr. Monroe introduces these into the feedback path, see, for example, Figure 12. In this case the compensation primarily affects the denominator of the transfer function (unless differentiators or high gain elements are used). Studies at Purdue have shown that it is also desirable to introduce independently terms into the numerator. This, of course, is more easily done by introducing the compensation in tandem rather than in the feedback path.

Another difficulty with this method is that since the artificial feel is produced in the feedback path the dual problems of producing satisfactory stability (with the compensation) and satisfactory artificial feel become interrelated, and consequently the solution is unduly complex.

In regard to the use of differentiators, the system may become oversusceptible to corrupting signals such as gusts. The double differentiation to obtain $-d^2\theta/dt^2$ may introduce undesirable responses when acting on sharply changing signals such as those from gusts.

In the yawing control system, Mr. Monroe fed back signals proportional to derivatives of the aerodynamic yawing angle β. These were obtained from the geometric yaw angle ψ by a relatively complex computor. Since he found that derivatives of the geometric pitch angle θ gave satisfactory stability improvement in the longitudinal control case, it is not quite clear why he did not try derivatives of ψ and simplify the computor.

Finally, the writers are not sure that the case for the use of an electric system is convincing. Mr. Monroe points out some of the difficulties with a mechanical system, but it is not believed these are severe. If the control system is dependent on two power systems (electric and hydraulic) it must be less reliable than if dependent on only one. It is believed that any system which is compensated electrically can be compensated equally well by hydraulic or mechanical elements.

Although these comments appear to be critical of Mr. Monroe's approach, they actually represent only differences of opinion concerning the detail design. The basic concept of introducing compensating elements into the system to improve its performance is of a great deal of importance to the field of airplane aerodynamics.

J. B. Rea (J. B. Rea Company, Inc., Los Angeles, Calif.): Mr. Monroe has performed a valuable service in presenting some of the basic principles of automatic control of aircraft in a manner designed for presentation to engineers not familiar with the field of aircraft stability and control. There has existed for some years a need for the interpretation and simple explanation of how and why an aircraft flies and of the methods for synthesizing automatic control devices to improve the performance of high-speed aircraft. It is to be hoped that the author will follow this first article with more detailed discussions of the various control devices which he has mentioned.

The device shown in Figure 17 would appear to have some disadvantages aside from the obvious advantages. The size of the backlash deliberately introduced would be chosen on the basis of two opposing considerations. The first is that the backlash should be small in order that operation under emergency conditions be satisfactory from the pilots' standpoint. On the other hand, the backlash should be large so that the stick motion is not velocity-limited during normal operations. A serious drawback to this system is that if the two autosyn pickups should become misaligned by an amount greater than the backlash, the system would then be completely in operation, since the pickups would operate in such a manner as to reintroduce error as fast as it is nulled out.

Assumedly, both the system presented in Figure 17 and the simpler one presented in Figure 16 are intended for use with autopilot systems. This would be accomplished in both cases by introducing autopilot control signals into the amplifier shown. These signals would act in the same manner as the misalignment previously mentioned. Thus in the system shown in Figure 16, autopilot signals would cause the control surface to move without any accompanying stick motion; whereas in the arrangement shown in Figure 17, for pilot operation of the autopilot pedestal controller, the maximum autopilot output signal must be on greater than the equivalent backlash. If this were not so, the control system would operate as if there were no surface position feedback; since the signals generated by the two autosyns would partially cancel resulting in a unidirectional control surface motion until the autopilot loop called for a reversal. This effective elimination of the control surface feedback probably would result in instability.

Steven E. Belsley and **William M. Kauffman** (National Advisory Committee for Aeronautics, Moffett Field, Calif.): Mr. Monroe has presented an interesting discussion which indicates the similarity both in equipment and in design methods between modern airplane control servomechanisms and those used in conventional land-based applications. The author has confined his examples to what might be termed "pilot-assist" devices, which alter the response and control characteristics of the airplane but still leave the pilot in normal continual control. In contrast to the usual completely automatic control system, rational design of these pilot-assist devices requires knowledge of the interrelation between the response and control characteristics of the human and of the airplane. This problem has been partially met to date by elaborate airplane handling-qualities specifications stemming from the pioneer work of R. R. Gilruth;[1] such specifications attempt to define regions of satisfactory airplane behavior on the basis of flight tests of a wide variety of airplanes.

With the advent of powerful pilot-assist devices such as control boosters and oscillation dampers, the airplane behavior can be varied readily over extreme ranges. Hence, information on the optimum and (from fail-safe considerations) minimum safe basic airplane characteristics becomes of great interest to the designer. Mr. Monroe mentions one increasingly popular method of attacking this problem; the use of "a mockup of the actual control surface system with control surface position presented to an analogue computer which determines the airplane response." Another approach used extensively by the National Advisory Committee for Aeronautics (NACA) Ames Aeronautical Laboratory in the past few years has involved flight tests of airplanes whose response and control characteristics can be varied in flight by means of servomechanisms. Such research equipment makes it possible to obtain in actual flight pilots' opinions and recorded data on the effects of various airplane characteristics which in production airplanes are amenable to change by pilot-assist devices. In addition to furnishing information on desired airplane and system characteristics, these research studies have given valuable experience in the design of satisfactory pilot-assist devices. One early example, an all electric servomechanism for artificially varying dihedral effect in flight, has been described in an NACA publication[2] along with the results of ensuing flight tests. Similar equipment to vary the airplane directional characteristics through servomechanism actuation of the rudder has been developed and successfully employed in later flight tests.

Another item of interest to the designer of airborne servomechanisms is the accuracy to which it is possible to simulate, and hence study, the flight behavior of proposed airplane-servomechanism combinations by means of bench tests, mockups, and analogue computers. This ability to simulate must be determined by comparison of the simulator results with those of carefully controlled flight tests of the airplane-servomechanism combination. An example study of this type is reported in two NACA publications.[3,4]

It might be noted that in addition to providing adequate simulation of the overall system, it is equally important in ground studies of proposed systems to ask the simulated equipment the proper questions; that is, to give the simulator problems which adequately cover the expected flight operating conditions. For example, Ames Aeronautical Laboratory developed and flight tested a force servomechanism similar to that shown in Figure 13, in which a control force proportional to $d^2\theta/dt^2$ is fed back automatically to the pilot's control stick. The system appeared to behave satisfactorily during ground mockup tests. However, in flight an instability developed when the gain was increased to provide adequate force feedback. Prior to actual instability the pilot complained of a "rattiness" in the stick force. In subsequent studies on a high-speed electronic simulator, this condition did not develop when specified control stick motions were applied; however, when a stick force input was used the same instability was readily apparent. The latter condition corresponds more closely to actual piloting procedure, in which the pilot is primarily influenced by the force variation and does not rigidly position the stick.

Mr. Monroe has chosen in Figure 8 to generate sideslip signals in an indirect fashion. In cases where it is feasible to provide suitable external mounting and to perform the necessary calibrations, it is much simpler to measure the angle of sideslip directly by a relative-wind direction pickup, such as a vane.

REFERENCES

1. REQUIREMENTS FOR SATISFACTORY FLYING QUALITIES OF AIRPLANES, **R. R. Gilruth.** *Technical Report 755*, National Advisory Committee for Aeronautics (Washington, D. C.), 1943.

2. AN APPARATUS FOR VARYING EFFECTIVE DIHEDRAL IN FLIGHT WITH APPLICATION TO A STUDY OF TOLERABLE DIHEDRAL ON A CONVENTIONAL FIGHTER AIRPLANE, **William M. Kauffman, Charles J. Liddell, Jr., G. Allan Smith, Rudolph D. Van Dyke, Jr.** *Technical Report 948*, National Advisory Committee for Aeronautics (Washington, D. C.), 1949.

3. PRACTICAL METHODS OF CALCULATION INVOLVED IN THE EXPERIMENTAL STUDY OF AN AUTOPILOT AND THE AUTOPILOT-AIRCRAFT COMBINATION, **Louis H. Smaus, Elwood C. Stewart.** *Technical Note 2373*, National Advisory Committee for Aeronautics (Washington, D. C.), 1951.

4. A COMPARISON OF PREDICTED AND EXPERIMENTALLY DETERMINED LONGITUDINAL DYNAMIC RESPONSES OF A STABILIZED AIRPLANE, **Louis H. Smaus, Marvin R. Gore, Merele G. Waugh.** *Technical Note 2578*, National Advisory Committee for Aeronautics (Washington, D. C.), 1951.

William R. Monroe: In this paper an attempt was made to discuss in a few pages the field of changing the dynamic response of airplanes by means of electric equipment. Of necessity this required only that a minimum of aerodynamic, electronics, and servomechanism theory be introduced. However, the author's feeling of possibly having given a misleading impression of unattainable simplicity of a complex subject has been dispelled by the excellent discussions which will correct any such impression that the reader might have gained. In particular Mr. Pines mentions the gain to be made by preliminary simplified analysis and the pitfalls of neglecting effects of system nonlinearity. In this regard it is believed that whenever possible actual system mockups with production parts be tested as a final phase of a development program. The problem of including a human pilot in a simulator control loop is extremely difficult, as was indicated by Mr. Pines, Dr. Clauser and Dr. Truxal, and Mr. Belsley and Mr. Kauffman. This problem is partially solved by making visual and control column force feedback paths to the pilot as realistic as possible. However, the author does not know of any successful way in which realistic time histories of airplane acceleration have been applied to the body of a pilot flying a simulator. The NACA research flight test projects to which Mr. Belsley refers are considered of extreme importance since this work goes a long way towards filling the gap between flight simulator work and airplane production engineering.

In answer to the remarks on the difficulty of using β instead of ψ for increasing the airplane stability, β was chosen for the purpose of indicating the possible use of electronic computers in airplane installations and does not constitute a recommendation for actual use of this particular system on an airplane. In addition to this it should be emphasized that the circuit shown for the β computer is intended only as an indication and, as pointed out by Dr. Clauser, should be improved and developed considerably if a β computer were to be used in an airplane. The use of pure differentiators and especially double differentiators is not recommended because of the high amplification of noise.

Those points mentioned by Mr. Gellert which were typographical errors occurring in the original manuscript have been corrected in the published text. The author is in agreement with Mr. Gellert's other remarks. In practice it is mandatory that the variation of the aerodynamic coefficients with air speed and altitude be considered. This matter was not discussed in the paper because of space restrictions. Figure 2 of the paper was used as a visual aid in discussing, perhaps in too simple a manner, the way an airplane is controlled. As in many simplified discussions the curves are not rigorous. For example, as Mr. Gellert points out, the curve showing θ as a function of M_T actually represents a dynamic relationship based on a differential equation and thus there are an infinite number of curves that could have been used for the illustration.

As indicated by Dr. Rea, the electric systems of Figures 16 and 17 of the paper present some practical difficulties that may become extremely important if such a system were carried beyond the idea stage and developed for airplane use. The amount of feedback introduced deliberately must be given careful consideration although it does not appear to be a particularly serious problem. In Figure 17 the hydraulic servo must be capable of moving the cable system as fast as the maximum rate at which the human pilot moves the control column if the backlash is to be effective. Otherwise the pilot when moving the control column probably would feel an objectionable bump as the backlash is suddenly taken up. On the other hand, the effect of too much backlash in emergency operation can be reduced considerably by using an electrically controlled clutch which removes the backlash when the normal electric servo control system fails.

In conclusion, the author would like to express agreement with Dr. Clauser that the case of the use of electric systems in primary airplane control systems is not by any means conclusive. However, it is believed that this application of electric equipment will have an important future provided that the problem of fail-safety and dependability is solved.

The Use of Nonlinear Feedback to Improve the Transient Response of a Servomechanism

J. B. LEWIS
ASSOCIATE MEMBER AIEE

IN THE design of a positional servomechanism, the desired result is a system whose output member will follow exactly the instantaneous position of the input member. This, of course, imposes requirements on the system when in the dynamic or transient state which cannot be realized.

One common test of the stability and fidelity of a linear servomechanism is its response to a step function. In the event that the relation between input and output positions of the servomechanism can be described by a second-order linear differential equation with constant coefficients, the transient response can be found rather simply. Nondimensional curves are usually plotted which describe the response of this type of system in terms of the two parameters ω_n, the natural angular velocity, and ζ, the damping ratio. It can be seen readily from an examination of these curves that, in general, the output responds with a high velocity for smaller values of the damping ratio but moves beyond the desired steady-state position before coming to rest. However, for large values of damping ratio, the output response is slower, but there is no overshoot. Thus the system designer must choose a suitable damping ratio which is a compromise between velocity of response and amount of overshoot. These same properties apply in a large measure to systems described by higher order linear differential equations since one mode of oscillation predominates.[1]

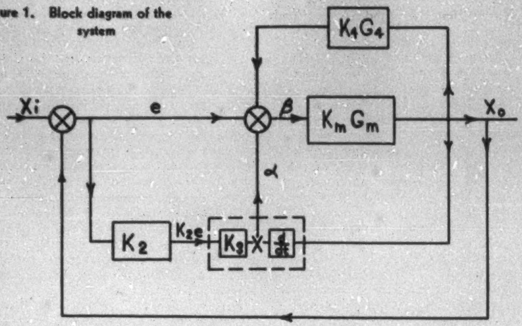
It is the purpose of this paper to investigate the possible improvement in the transient response of a second-order linear positional servomechanism that results from the introduction of a type of feedback which causes the damping ratio to vary during the period of the transient. The feedback is so arranged that the system has a small damping ratio during the early part of the transient response and thus a high initial velocity results. As the output member approaches the steady-state position, however, the damping ratio becomes large and the overshoot is minimized or entirely eliminated. This type of feedback can be accomplished by the multiplication of two variables in the feedback path instead of the usual addition or subtraction used in linear servomechanisms.

A block diagram of the system is shown in Figure 1. The block marked X indicates the multiplication in a tachometer generator in the feedback path producing the nonlinearity and the K_4G_4 block represents a second tachometer generator providing conventional tachometric feedback. The amplified error signal is used to excite the field of the tachometer in X, since it is desired to have a large output voltage from the tachometer generator during the early part of the transient and a small output voltage as the system reaches its steady-state position. In this way, the damping ratio is made small when the error is large if the armatures of the two tachometers are connected in series so that their output voltages subtract. The damping then increases as the error decreases since the field current

Paper 52-199, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952, and approved by the AIEE Committee on Technical Operations for representation at the AIEE Middle Eastern District Meeting, Toledo, Ohio, October 28–30, 1952. Manuscript submitted February 11, 1952; made available for printing September 5, 1952.

J. B. Lewis is with the University of Tennessee, Knoxville, Tenn.

Figure 1. Block diagram of the system



of the K_4G_4 tachometer is constant. K_mG_m is the transfer function of the d-c output motor. The equations associated with the block diagram are

$$e = x_i - x_o, \quad \alpha = K_2K_3e\frac{dx_o}{dt}, \quad \text{and}$$

$$\beta = e + K_2K_3e\frac{dx_o}{dt} - K_4\frac{dx_o}{dt}$$

The schematic circuit diagram of the system is shown in Figure 2, where K_1 and K_2 are balanced d-c amplifiers using amplidynes in the output stage. The differential equation relating output and input is

$$J'\frac{d^2x_o}{dt^2} + \left(f' + \frac{K_1K_4K_5K_TI_t}{R_a} + \frac{K_7K_T}{R_a}\right)\frac{dx_o}{dt} - \left(\frac{K_3K_1K_5K_2K_TE}{R_fR_aX_m}\right)(x_i - x_o)\frac{dx_o}{dt} + \frac{EK_TK_1K_6}{X_mR_a}x_o = \frac{EK_TK_1K_6}{X_mR_a}x_i$$

provided L_a and L_f are neglected. This is a second-order nonlinear differential equation. The third term in the differential equation is the one introduced by the nonlinear feedback. If $K_2 = 0$, this term vanishes and the equation reduces to a second-order linear differential equation with constant coefficients.

To simplify the writing of this equation, it is convenient to define the following constants

$$A = f' + \frac{K_4K_1K_5K_TI_t}{R_a} + \frac{K_7K_T}{R_a},$$

$$B = \frac{K_3K_1K_5K_2K_TE}{R_fR_aX_m}, \quad D = \frac{EK_TK_1K_6}{X_mR_a}$$

The equation becomes

$$J'\frac{d^2x_o}{dt^2} + A\frac{dx_o}{dt} - B(x_i - x_o)\frac{dx_o}{dt} + Dx_o = Dx_i$$

Since this investigation is to be limited to finding the response of the system to a step input, the input position x_i is written as $|x_i|$ to indicate that it is a constant for all positive values of time. The differential equation then is

$$J'\frac{d^2x_o}{dt^2} + (A - B|x_i|)\frac{dx_o}{dt} + Bx_o\frac{dx_o}{dt} + Dx_o = D|x_i|$$

It is an equation of the form: $y'' + ay' + byy' + cy = d$. Because no satisfactory method of solution could be found, it appeared logical to carry out a step-by-step (sometimes called point-by-point) solution since a result can be obtained and the accuracy is largely dependent only on the number of points which are calculated. It also has the advantage of retaining the physical significance of the problem.

The original differential equation may be rewritten

$$J'\frac{d^2x_o}{dt^2} + (A - B|x_i| + Bx_o)\frac{dx_o}{dt} + Dx_o = D|x_i|$$

It is then convenient to let $R = A - B|x_i| + Bx_o$ and $G = A - B|x_i|$, so that $R = G + Bx_o$. R is a linear function of x_o and thus increases during the transient response from G, at the first instant when the step input is applied, to A when the system reaches its steady-state position where $x_o = |x_i|$. If, as an approximation, R is regarded as a constant over a small increment of time, the system can then be treated as being linear (that is, representable by a linear differential equation) over this increment. In successive increments of time R becomes larger, and therefore the actual system is approximated as a series of linear systems with increasing damping ratios as the output moves from its initial to its final position.

Let $R = R_n$ during the nth increment of time. Then during this nth increment the differential equation becomes:

$$J'\frac{d^2x_o}{dt^2} + R_n\frac{dx_o}{dt} + Dx_o = D|x_i|$$

which is a linear differential equation with constant coefficients. The characteristic equation is $J'm^2 + R_nm + D = 0$ and its roots are

$$m_1, m_2 = -\frac{R_n}{2J'} \pm \sqrt{\frac{R_n^2}{4J'^2} - \frac{D}{J'}}$$

Now if we let $\omega_0 = \sqrt{D/J'}$ and $R_c = 2\sqrt{J'D}$, and $\zeta_n = R_n/R_c$, then the roots are $m_1, m_2 = -\zeta_n\omega_0 \pm j\omega_0\sqrt{1 - \zeta_n^2}$ where $j = \sqrt{-1}$ and $\zeta_n < 1$, or finally

$$m_1, m_2 = -\zeta_n\omega_0 + j\omega \text{ for } \omega = \omega_0\sqrt{1 - \zeta_n^2}$$

The general solution for x_o during the nth increment is

$$x_o = x_{oss} + K_n\epsilon^{-\zeta_n\tau}\sin(\sqrt{1 - \zeta_n^2}\tau + \phi_n)$$

where $x_{oss} = |x_i|$, $\tau = \omega_0t$, and K_n and ϕ_n are constants to be evaluated from the

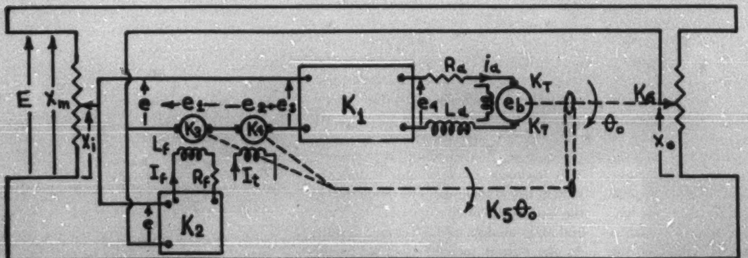


Figure 2. Schematic circuit diagram

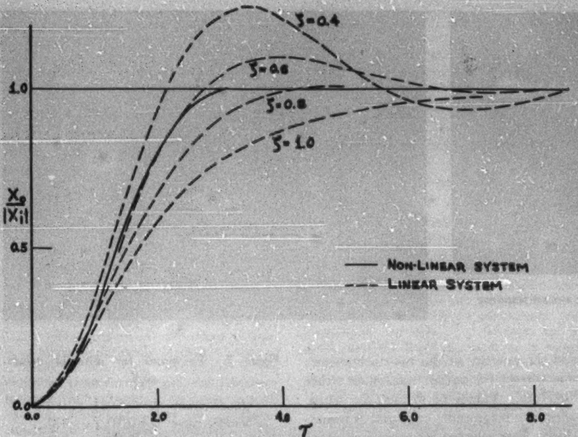


Figure 3. Comparison of second-order system with variable damping ratio with second-order linear system

initial conditions of the nth increment. These initial conditions are actually the output position (x_o) and velocity (dx_o/dt) at the end of the $(n-1)$th increment of time since these quantities must be continuous. The transient response of the servomechanism can now be investigated in terms of ω_0, which is a constant, and ζ. ζ is a discontinuous function of x_o since $\zeta_n = R_n/R_c = G/R_c + B/R_c\ x_o$, and R_n is constant during the nth increment. Therefore, x_o in this expression should properly be called x_{on}, a constant during the nth increment, and it may be taken, as a first approximation, as x_o at the beginning of the nth increment.

It is interesting to consider the variation in ζ in more detail. In the limit, as the length of the time increments approaches zero, ζ becomes a continuous linear function of x_o. This may be written as $\zeta = \zeta_0 + sx_o$ where ζ_0 is ζ at the instant the step input is applied and s is the slope of the $\zeta(x_o)$ line. Therefore $\zeta_0 = G/(2\sqrt{J'D}) = (A - B|x_i|)/(2\sqrt{J'D})$ and $s = B/(2\sqrt{J'D})$ from the previous definitions of R and R_c. As the output reaches its steady-state position, ζ reaches a final value which can be called ζ_∞ and $\zeta_\infty = \zeta_0 + s|x_i|$ since $x_o = |x_i|$ in the steady state. Thus the values of ζ_0 and ζ_∞ fix the manner in which the damping ratio ζ varies during the transient.

To obtain a result by this method, values of $\zeta_0 = 0.1$ and $\zeta_\infty = 1$ were chosen as being typical values which can be achieved in an actual system. The solution was carried out in terms of τ instead of actual time for greater generality. A sample calculation is included in Appendix I to illustrate the method. The increments of τ were chosen so the variation in ζ between adjacent increments was never greater than 0.05. One refinement was included in the calculation in order to increase the accuracy. As a first approximation, the variation in x_o was calculated in each increment with ζ_n determined by x_{on}, the value at the beginning of the increment. Then after the value of x_o at the end of the increment (x_{on+1}) was calculated, the average of x_o and x_{on+1} was used to determine a new value of ζ_n as a second approximation. A second solution for x_o in this same increment was then used to calculate a second value of x_{on+1} to be used in turn, in the calculation for the following increment. The solution for x_o in any increment was adjusted for the initial conditions of each increment in both the first and second approximations.

A plot of x_o versus τ is shown in Figure 3. In order to make a comparison with the transient response of the nonlinear system, transient response curves of a second-order linear system for values of $\zeta = 0.4$, 0.6, 0.8, and 1.0 also are included in Figure 3. It can be seen that for the particular values of ζ_0 and ζ_∞ chosen for the calculation, the nonlinear system responds with a velocity very nearly that of a linear system with a damping ratio of 0.6. However, in the case of the nonlinear system there is no overshoot.

To visualize the operation of the nonlinear system as approximated in terms

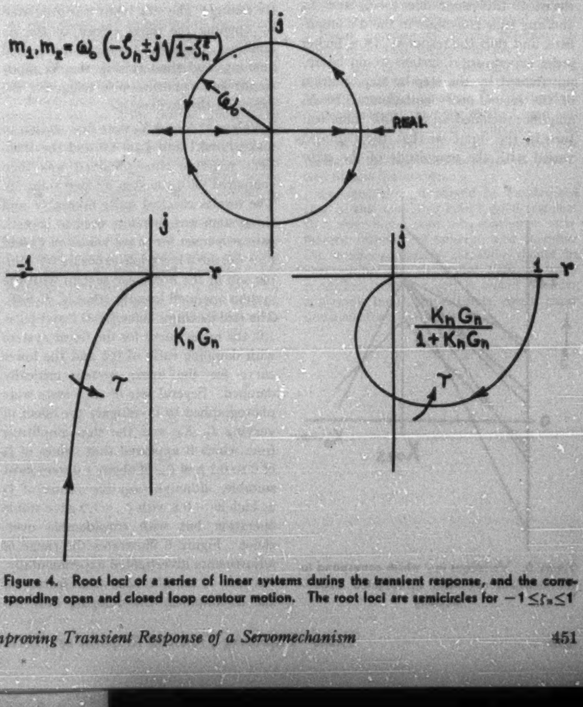

$m_1, m_2 = \omega_0\left(-\zeta_n \pm j\sqrt{1-\zeta_n^2}\right)$
j
ω_0
REAL
j
-1
r
K_nG_n
τ
j
1
r
$\frac{K_nG_n}{1+K_nG_n}$
τ

Figure 4. Root loci of a series of linear systems during the transient response, and the corresponding open and closed loop contour motion. The root loci are semicircles for $-1 \leq \zeta_n \leq 1$

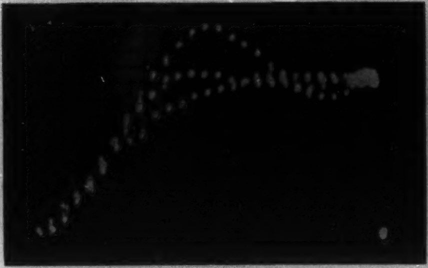


Figure 5. Comparison of linear and nonlinear system response

of familiar linear system KG contours, Figure 4 was included. It illustrates the root loci in the usual s-plane in terms of ζ_n and the corresponding K_nG_n and $(K_nG_n)/(1+K_nG_n)$ contours with the directions for increasing τ indicated.

To test the method for other combinations of ζ_0 and ζ_∞ an experimental investigation was considered advisable since the calculation of several curves would have been very tedious. The system which was built had an ω_0 of 3.97 radians per second, giving a free period of 1.58 seconds. This was considered slow enough to warrant neglecting the electrical transients due to L_f and L_a and any time constants in the d-c amplifiers, and thus the response of the higher order experimental system would be approximated by the step-by-step solution of the second-order equation. The K_2 amplifier included an absolute value feature in the input so that the damping varied with the magnitude of the error and the polarity of the two tachometers was correct for output motion in either direction. Values of I_t and K_2, which affect the magnitudes of A and B respectively, were selected for experimental investigation since A and B fix ζ_0 and ζ_∞ independently after the other system constants and $|x_i|$ are chosen.

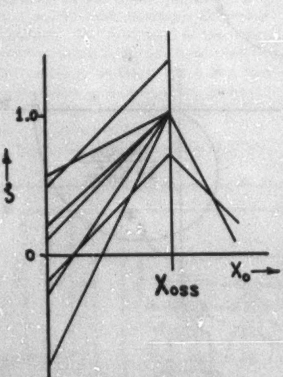


Figure 6. Variations in ζ which correspond to the transient responses investigated experimentally

The transient responses were recorded by photographing the screen of an oscilloscope where a voltage proportional to output position was applied to the vertical plates and a driven linear sweep was used on the horizontal plates. Intensity modulation of 16.3 cycles per second was used for timing. The step input was simulated by opening the input circuit to the K_1 amplifier, displacing the input the desired amount, and then closing the K_1 input circuit and simultaneously triggering the sweep on the oscilloscope.

Values of I_t and K_2 were first chosen to make $\zeta_0=0.1$ and $\zeta_\infty=1.0$ and the transient response thus obtained was then compared with the step-by-step solution. The results checked quite favorably and the system was therefore used to investigate responses for other values of ζ_0 and ζ_∞. Figure 5 is a triple exposure for comparison of the nonlinear system with the system operated linearly (that is, $K_2=0$). The middle curve is for $\zeta_0=0.1$ and $\zeta_\infty=1.0$, the upper curve for the linear system with damping ratio of 0.4 and the lower curve for the linear system critically damped. Several sets of transients were photographed to investigate the effect of varying I_t, K_2, and the step amplitude from which it appeared that values of ζ_0 of 0 to 0.1 and ζ_∞ of about 1.0 were most suitable, although negative values of ζ_0 as high as -0.8 with $\zeta_\infty=1.0$ gave stable operation but with considerable overshoot. Figure 6 illustrates the range of adjustments investigated experimentally. Because of the absolute value feature of the K_2 amplifier, the damping ratio decreases if there is any overshoot $(x_o>|x_i|)$ and thus a stability limit is imposed on the system. If the overshoot is sufficient to give negative values of ζ, the system is unstable. The absolute value feature could, however, be eliminated if operation in one direction only is desired. Figure 7 illustrates the effect of using different values of K_2.

Figure 7. Responses for different values of K_2

Upper curve: $\zeta_0=-0.84$, $\zeta_\infty=1.0$
Middle curve: $\zeta_0=0.10$, $\zeta_\infty=1.0$
Lower curve: $\zeta_0=0.20$, $\zeta_\infty=1.0$

It thus appears that for systems subjected to step inputs which can be approximated as second order, some improvement in the transient response is possible using the type of feedback described. More complete results for other inputs and for higher order systems could be more advantageously investigated using an analogue computer.

Appendix I. Sample Calculation by Step-by-Step Method

1. For $\zeta_0=0.1$

$x_o=1+K_1\epsilon^{-4.1\tau}\sin(0.995\tau+\phi_1)$

2. At $\tau=0$, $x_o=0$ and $\dfrac{dx_o}{dt}=0$

therefore

$0=1+K_1\sin\phi_1$

and

$0=-0.1\sin\phi_1+0.995\cos\phi_1$

from which

$\tan\phi_1=9.95$ and $\phi_1=84.3$ degrees

Also

$K_1=-1/\sin\phi_1=-1.005$

3. Finally $x_o=1-1.005\epsilon^{-0.1\tau}\sin(0.995\tau+$

84.3 degrees) during the first increment.

4. If $\tau=0.2$, $x_o=0.0100$.

5. The average x_o during the first increment, $x_{oav}=0.0050$, from which the average ζ during this increment may be found.

$$\begin{aligned}\zeta_{av}&=\zeta_0+sx_{oav}\\&=0.1+0.9\times0.0050\\&=0.1045\end{aligned}$$

6. The revised expression for x_o becomes by repeating step 2

$$x_o=1-1.005\epsilon^{-0.1045\tau}\sin(0.994\tau+84.0\text{ degrees})$$

7. If $\tau=0.2$, $x_o'=0.0201$ and $\dfrac{dx_o'}{dt}=0.1944$.

8. The values from step 7 are then used to calculate the response during the next increment of τ.

Reference

1. Principles of Servomechanisms (book), G. S. Brown, D. P. Campbell. John Wiley and Sons, Inc., New York, N. Y., 1948, page 53.

Discussion

David B. Stallard (Westinghouse Electric Corporation, Baltimore, Md.): This paper is an excellent example of how a given servomotor in a positional control system may be compensated by deliberately nonlinear techniques so as to produce a transient response decidedly superior to that of a purely linear system. In addition to the improved transient response, this nonlinear tachometric feedback can be made to yield another advantage, that is, increased steady-state positional accuracy. The actual positional accuracy of a second-order servo usually is limited by coulomb (speed-invariant) friction on the output member, gearing, and servomotor. Hence the accuracy is improved if the torque constant (ratio of output torque to steady-state error) is increased; the torque constant is important in accurate military computer servos and especially in servos subject to significant fluctuating load torques. If nonlinear tachometric feedback is applied to increase the damping with decreasing error, then the loop gain may be increased without causing overshoots, and both the torque constant and steady-state accuracy will be improved.

In addition to the ingenious step-by-step method of solution described in this paper, at least one other method can be used advantageously. The second-order differential equation in x_0 and x_i can be rearranged to solve for $(d^2x_0)/(dt^2)$. This acceleration can be calculated for the first instant of the transient, and then can be multiplied by a suitably short time interval, Δt, to yield the approximate velocity $\Delta dx_0/dt|_1$ at the end of the interval. Substitution of this velocity and the output angle Δx_{01} (essentially zero) into the differential equation yields a new acceleration, $d^2x_0/dt^2|_1$. Again, $d^2x_0/dt^2|_1(\Delta t)=\Delta dx_0/dt|_2$, and likewise, $dx_0/dt|_1\ (\Delta t)=\Delta x_{02}$. The acceleration $d^2x_0/dt^2|_2$ can be calculated from the differential equation, and the process can be repeated in tabular form until the whole transient is calculated.

The time interval Δt may be shortened if more accuracy is necessary. Clearly, this method is simple to learn, remember, and use; the calculations are straightforward arithmetic and no extra parameters are introduced. It also may be used for a more complex input x_i, such as a ramp input of constant velocity, such that $x_i=vt$; at the beginning of each new time interval, the instantaneous x_i also could be substituted into the differential equation.

In many nonlinear systems, the phase plane is quite useful for analysis; error velocity is plotted versus error on linear Cartesian co-ordinates. A second-order differential equation for error can be converted into an equation for e and de/dt; for instance, acceleration $d^2e/dt^2=de/dt\ d/de\ (de/dt)$. A transient plots as a trajectory in the phase plane terminating at the origin; if acceleration is constant, the trajectory is a parabola.

Another nonlinear feedback technique can be utilized advantageously also, with only one tachometer instead of two. The alternating-voltage signal from a drag-cup a-c tachometer can be applied to the control grid of a remote-cutoff pentode which is biased by a negative voltage proportional to the magnitude of the error voltage. In a suitably designed pentode amplifier the feedback gain would be almost inversely proportional to error within limits, the maximum gain (for small error) being about 20 times the minimum gain (for large errors). Because the damping would increase with decreasing error, the compensated servo would have a transient response similar to that described in the paper.

Still another technique would be to switch with a relay from one tachometer gain to a much larger gain when the error falls below a prescribed amount; the switching may be made dependent on error velocity as well as error. Such a system, where the differential equation of error changes abruptly, may be termed a multiple-mode system.

In the future, designers of servomechanisms for stringent specifications of performance, weight, and volume should seek to exploit the rich possibilities of nonlinear techniques and multiple-mode operation.

This paper by Professor Lewis represents a definite contribution to feedback control.

J. B. Lewis: The author wishes to acknowledge with graditude the discussion by Mr. Stallard. This discussion adds considerably to the value of the paper in pointing out some of the practical advantages to be gained in certain cases by using the type of feedback described.

The alternate step-by-step method suggested is certainly much simpler to apply and should yield accurate results if the time increments are sufficiently small. This method, and also the method used in the paper, can be employed to obtain solutions for other types of input provided the input can be approximated as a constant during each time increment, and a solution for other inputs certainly would be useful. However, the damping ratio would not be a linear function of output position for inputs other than a step function.

The use of an a-c tachometer and a remote-cutoff pentode is an interesting and useful suggestion and has the advantages of greater speed than the method used and the elimination of a d-c amplifier. The results should be the same provided the variation in amplification factor as a function of bias is suitable.

In conclusion, it should be emphasized that in this nonlinear case a good response to a step input does not guarantee satisfactory response in general, and therefore the nonlinear system certainly should be checked for the particular type of input conditions under which it would operate; a periodic input, for example, might cause unsatisfactory operation.

Automatic Synchronizing for Aircraft Alternators

FRED B. McCARTY
ASSOCIATE MEMBER AIEE

G. W. HILLS
ASSOCIATE MEMBER AIEE

THE PIONEERING of 3-phase 400-cycle constant frequency electric systems for modern aircraft has presented many difficult problems but has given a basic solution to the problem of meeting unprecedented electric load requirements. Load requirements have increased because of the increasing need for complex electric control systems and electronic devices in the maintenance of air supremacy. As a consequence of this greater dependence on electric power, greater reliability and simplicity of operation of the power systems are now demanded.

Therefore, to this end, an automatic synchronizing system has been designed. It simplifies the procedure of paralleling aircraft alternators by holding all alternators of the system in a state of synchronism, ready for paralleling at the operator's discretion. Synchronizing is accomplished by very accurately controlling the speed of each alternator by electrohydraulic means.

Need for Automatic Synchronizing of Aircraft Alternators

Four 40-kva 0.75-power factor 120-volt/208-volt 3-phase 400-cycle Westinghouse alternators supply the electric power for a Convair *B-36* aircraft. The alternators are driven through a Sundstrand constant-speed hydraulic drive from each of four reciprocating engines. Four units, rather than one, are used to insure continuity of electric power in the event of a reciprocating engine or alternator failure. Three of the four alternators are paralleled to provide greater stability under high-surge loads and also better fault-clearing capacity. The fourth alternator is kept in a stand-by condition as an emergency power source.

The manual operation of paralleling synchronous machines is familiar to all electrical engineers. Voltage and frequency of the incoming alternator are adjusted to be the same as that of the system; then the alternator circuit breaker is closed when the phase difference is small. This procedure is quite tedious. As a result, it is believed that synchronizing errors occur and that they have been responsible for some severe abuse to components and supporting structure of the constant-speed hydraulic drive and the alternator. R. P. Judkins and H. M. McConnell of Westinghouse Electric Corporation have shown that polarity reversal of the d-c exciter for the alternator field is an occasional result of out of phase paralleling. Polarity reversal of the exciter is not tolerable because the alternator terminal voltage will oscillate when the voltage regulator senses reversed exciter polarity.

Therefore, one of the reasons for developing an automatic synchronizer was to reduce the possibility of paralleling alternators out of phase. However, another primary reason favoring the development of automatic synchronizing is the trend in large aircraft design toward automaticity to permit small flight crews

General Description of the Automatic Synchronizing System

A question arose as to which would be the better solution to the problem: a completely automatic paralleling system which would bring the alternator into frequency and close the circuit breaker at the proper time, or an automatic synchronizing system to bring the alternator into phase with bus frequency and hold it in phase until the circuit breaker was manually closed. On the basis of several considerations—fail safe characteristics, simple manual override features, simplicity of design, and light weight—it was decided that the automatic synchronizing system was superior.

The system shown in Figure 1 is a very simple synchronizer which has been flight-tested at Convair, and despite its limitations it was enthusiastically received by flight engineers. The function of the system is to cause an incoming alternator to "lock in" to synchronism with the bus frequency whenever the alternator frequency is manually adjusted to within approximately 2 cycles of bus frequency. Hence the limitation of the system is that it requires some manual adjustment of the alternator frequency. Nevertheless, the tedious task of zero-beating the alternator and bus frequencies is performed automatically, which should substantially reduce paralleling errors.

The simplicity of the system is appreciated when it is considered that the total weight of the synchronizer is only 0.952 pounds per alternator, including wiring. The new parts required are a relay, a fuse, a resistor, and a capacitor. The synchronizer utilizes the existing General Electric load division circuit so that this circuit performs a dual function. When the alternator is paralleled with other alternators, the load division circuit performs its normal function. (The circuit's normal function is to sense phase *B* alternator current and compare it with phase *B* voltage so that a direct-current signal proportional to the real component of load current will be developed. This signal is compared with signals from other alternators in the system and will change the torque output of each alternator drive until equal load division exists.) But when the alternator circuit breaker is opened, an auxiliary switch on the circuit breaker operates a relay which switches the input signal of the load division circuit so that it performs the synchronizing function. Ordinarily, the load division circuit would have had no useful function when the alternator circuit breaker was opened.

The automatic synchronizing circuit shown schematically in Figure 2, and as a block diagram in Figure 3, is a refinement of the circuit of Figure 1. It includes a differential selsyn to equalize automatically alternator frequency with bus frequency; the synchronizer then locks in or zero-beats the alternator with bus frequency. Therefore, this system is entirely automatic except that circuit-breaker closure is left to the discretion of the flight engineer. There are several obvious methods for accomplishing automatic circuit breaker closure, but for safety reasons they have not been considered. This system has not been flight-

---

Paper 52-311, recommended by the AIEE Air Transportation Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Middle Eastern District Meeting, Toledo, Ohio, October 28–30, 1952. Manuscript submitted June 5, 1952; made available for printing September 5, 1952.

FRED B. McCARTY and G. W. HILLS are with the Consolidated Vultee Aircraft Corporation, Fort Worth, Tex.

The authors gratefully acknowledge the contributions of their Convair associates: E. C. Harmon, design engineer, who participated in the vector analysis of the synchronizing circuit which led to its development, who industriously contributed to preliminary tests, and who proposed the use of a capacitor as the most practical phase shifting element; F. R. Liggitt, design group engineer, who pointed out the need for automatic synchronizing and outlined the basic requirements along which this system was developed; and J. M. Murphy, senior design engineer, who helpfully criticized the electric circuitry of the synchronizer.

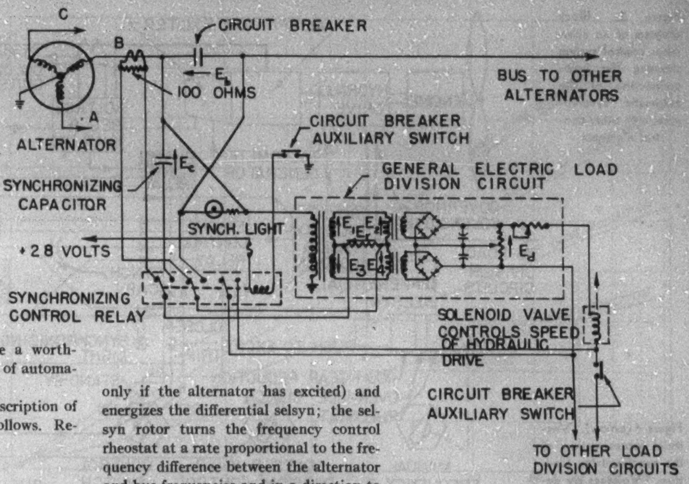

C
CIRCUIT BREAKER
B
E_b
100 OHMS
BUS TO OTHER ALTERNATORS
A
ALTERNATOR
CIRCUIT BREAKER AUXILIARY SWITCH
E_c
SYNCHRONIZING CAPACITOR
GENERAL ELECTRIC LOAD DIVISION CIRCUIT
SYNCH. LIGHT
E_1
E_2
E_3
E_4
E_d
+28 VOLTS
SYNCHRONIZING CONTROL RELAY
SOLENOID VALVE CONTROLS SPEED OF HYDRAULIC DRIVE
CIRCUIT BREAKER AUXILIARY SWITCH
TO OTHER LOAD DIVISION CIRCUITS

Figure 1. A simple alternator synchronizing system which has been flight-tested on a Convair B-36. Manual frequency adjustment is required if the initial frequency difference is greater than 2 cycles

tested but is expected to be a worthwhile solution to the problem of automatic synchronizing.

A step-by-step electrical description of the automatic synchronizer follows. Refer to Figures 2 and 3.

Control Switch "Off"

The alternator is de-excited and the circuit breaker is open.

Control Switch to "Stand-by"

The alternator can be excited; synchronizing control relay closes (will close only if the alternator has excited) and energizes the differential selsyn; the selsyn rotor turns the frequency control rheostat at a rate proportional to the frequency difference between the alternator and bus frequencies and in a direction to bring the frequency difference to zero. The frequency difference will be very small within 2 seconds, sufficiently small that lock-in to synchronism can be accomplished by the load division circuit.

The load division circuit electrically controls the hydraulic constant-speed drive and will change the speed of the drive until the difference voltage across the circuit breaker is very low. The synchronizing light will then go out and stay out so long as no large disturbances occur

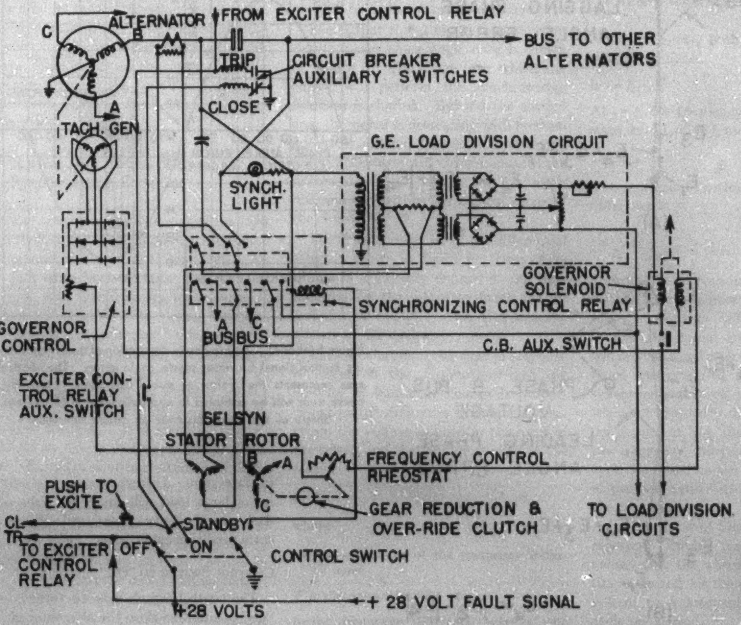

C
ALTERNATOR
B
A
FROM EXCITER CONTROL RELAY
BUS TO OTHER ALTERNATORS
TRIP
CLOSE
CIRCUIT BREAKER AUXILIARY SWITCHES
TACH. GEN.
G.E. LOAD DIVISION CIRCUIT
SYNCH. LIGHT
GOVERNOR SOLENOID
SYNCHRONIZING CONTROL RELAY
GOVERNOR CONTROL
A BUS
C BUS
C.B. AUX. SWITCH
EXCITER CONTROL RELAY AUX. SWITCH
SELSYN
STATOR
ROTOR
B
A
C
FREQUENCY CONTROL RHEOSTAT
GEAR REDUCTION & OVER-RIDE CLUTCH
PUSH TO EXCITE
CL
TR
STANDBY
ON
OFF
TO EXCITER CONTROL RELAY
CONTROL SWITCH
TO LOAD DIVISION CIRCUITS
+28 VOLTS
+ 28 VOLT FAULT SIGNAL

Figure 2. An automatic alternator synchronizing system which is not limited by the initial frequency difference

Figure 3. Block diagram of an alternator control system showing the interconnection of the automatic synchronizer with other control elements

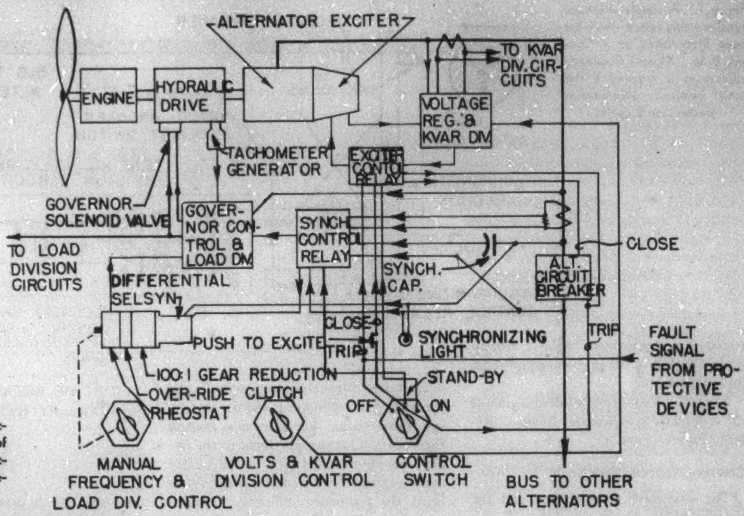
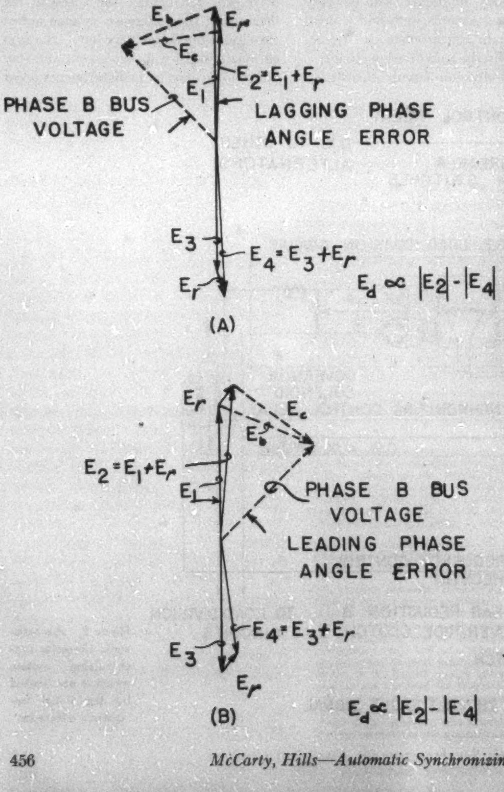

ALTERNATOR EXCITER
TO KVAR DIV. CIRCUITS
ENGINE
HYDRAULIC DRIVE
VOLTAGE REG. & KVAR DIV
TACHOMETER GENERATOR
EXCITER CONTROL RELAY
GOVERNOR SOLENOID VALVE
GOVERNOR CONTROL & LOAD DIV
SYNCH CONTROL RELAY
CLOSE
TO LOAD DIVISION CIRCUITS
ALT. CIRCUIT BREAKER
DIFFERENTIAL SELSYN
SYNCH. CAP.
CLOSE
SYNCHRONIZING LIGHT
PUSH TO EXCITE
TRIP
TRIP
FAULT SIGNAL FROM PROTECTIVE DEVICES
100:1 GEAR REDUCTION
OVER-RIDE CLUTCH
STAND-BY
RHEOSTAT
OFF
ON
MANUAL FREQUENCY & LOAD DIV. CONTROL
VOLTS & KVAR DIVISION CONTROL
CONTROL SWITCH
BUS TO OTHER ALTERNATORS

Figure 4 (below). Vectorial representation of the synchronizing function. Voltages are defined in Figure 1


E_b
E_r
E_c
$E_2 = E_1 + E_r$
E_1
PHASE B BUS VOLTAGE
LAGGING PHASE ANGLE ERROR
E_3
$E_4 = E_3 + E_r$
E_r
$E_d \propto |E_2| - |E_4|$
(A)
E_r
E_c
E_b
$E_2 = E_1 + E_r$
E_1
PHASE B BUS VOLTAGE
LEADING PHASE ANGLE ERROR
$E_4 = E_3 + E_r$
E_3
E_r
(B)
$E_d \propto |E_2| - |E_4|$

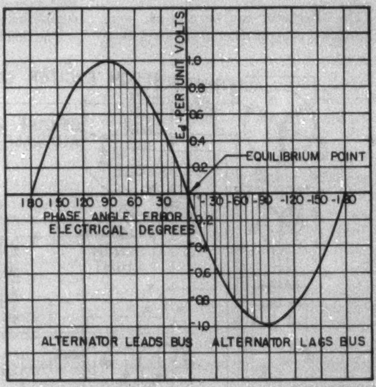


Figure 5 (above). Approximate description of the synchronizing control signal E_d versus phase angle error. The shaded area represents the region of stable operation. Normally, phase error will be restricted to approximately ±20 degrees. Shape of the curve depends on circuit constants

in the system. Disturbances such as adding a large load will cause the light to flash on and off a few times until a steady-state condition again exists.

Control Switch to "On"

The control switch may be turned to "on" position to close the alternator cir-

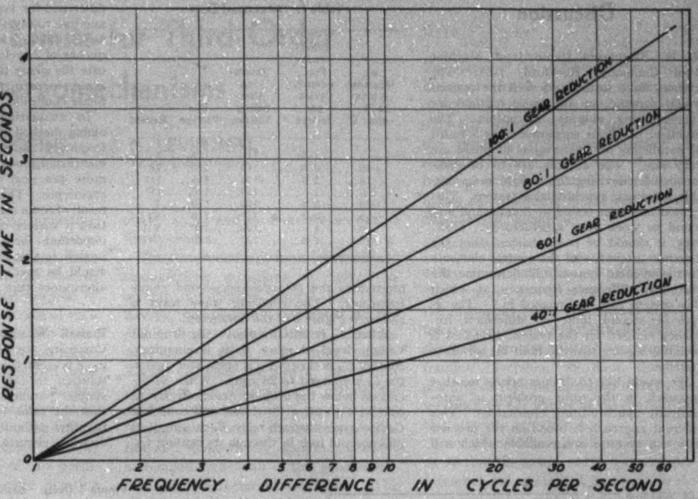


Figure 6. Response time versus frequency difference for various speed ratios of the differential selsyn

cuit breaker whenever the indicator light is off, as that indicates zero or low voltage across the circuit breaker.

In the on position, the synchronizing control relay is opened, which switches the input of the load division circuit to the current transformer, thus restoring the circuit to its normal function. The 100-ohm resistor across the current transformer prevents the occurrence of high-transient voltages during the switching operation.

Electrical Description of the Synchronizing Function

The synchronizing function of the load division circuit is explained by the vector diagrams of Figure 4(*A*) and 4(*B*). An elementary schematic of the load division circuit with synchronizing input signal is shown in Figure 1 to define voltages. $\mathbf{E}_1$ and $\mathbf{E}_3$ are reference voltages which are equal and opposite, $\mathbf{E}_1$ being in phase with phase *B* alternator voltage. To each of these voltages is added $\mathbf{E}_r$, which is a small component of $\mathbf{E}_b$, the other component being $\mathbf{E}_c$. It is essential that $\mathbf{E}_r$ should be nearly 90 degrees out of phase with $\mathbf{E}_b$ and of small magnitude, and this is insured by choosing a proper value for the capacitor (0.5 microfarad). The vector sum of $\mathbf{E}_1+\mathbf{E}_r$ is $\mathbf{E}_2$, and the vector sum of $\mathbf{E}_3+\mathbf{E}_r$ is $\mathbf{E}_4$. $\mathbf{E}_2$ and $\mathbf{E}_4$ are separately transformed to two rectifier circuits in such a way that the direct voltage output of the load division circuit E_d will be proportional to the difference of the absolute values $|E_2|$ and $|E_4|$. Also, the polarity of E_d will depend on whether the alternator voltage lags or leads bus voltage. Hence E_d will cause the hydraulic constant speed drive momentarily to increase or decrease speed to bring the phase error to zero.

The output signal of the load division circuit versus phase angle error is shown in Figure 5. The range over which the circuit will perform the synchronizing function is shaded. Beyond the shaded area the signal is decreasing with increasing error. This is an unstable condition and synchronism will not obtain until the differential selsyn brings the frequency difference to within the control range of the circuit.

The time of response of the differential selsyn is shown in Figure 6. Time of response is arbitrarily defined as the time required for the selsyn to bring the alternator to within 1 cycle of a bus frequency of 400 cycles.

The calculation of the time of response is based on the assumption that the velocity of frequency change is proportional to the difference in frequency between the voltage applied to the rotor of the selsyn and the voltage applied to the stator.

$$v=Kf$$

where

v = rate of change of the frequency difference

K is the proportionality constant

f = the frequency difference between the alternator voltage and the bus voltage.

Since time is the unknown value, the equation is set up in differential form

$$\frac{df}{dt}=Kf \text{ or } \frac{df}{f}=Kdt$$

After integrating both sides of the equation, it becomes

$$\ln f=Kt+C$$

At $t=o$, the frequency difference will be some initial value f_o and

$$C=\ln f_o$$

The expression for time at any frequency difference is

$$t=\frac{1}{K}\ \ln\frac{f}{f_o}$$

From this expression it can be seen that infinite time is required for the system to reach a frequency difference of zero, although the actual frequency difference will be very small after only a few seconds. Since the proposed system provides a synchronizing force which is effective when the frequency difference is 2 cycles per second or less, the time required for the frequency difference to become one cycle was used as the basis for selection of the gear ratio between the selsyn and the control rheostat.

In the curves of Figure 6 the ordinates represent initial frequency difference values and the abscissas represent the time required for the selsyn to drive the system to a frequency difference of 1 cycle per second.

Discussion

R. H. Eisengrein (Sundstrand Machine Tool Company, Rockford, Ill.): The authors have stated very well the reasons which demand an automatic method for synchronizing aircraft alternators. The practical problems encountered as a result of paralleling alternators out of phase are very real. The authors are to be commended for devising this unique and workable automatic synchronizing system. This system has been tested at Sundstrand and found to work very satisfactorily. However, it should be re-emphasized that the dynamic properties of the control elements in a particular system will determine the maximum difference frequency at which this system can be "phased in." The 2-cycle-per-second limit on difference frequency realized for the system discussed is probably a very realistic limit for all other systems.

We would like to discuss briefly another approach to the same problem of automatically synchronizing alternators. This different approach is based on the premise that two systems are available which will be preset to 400 cycles per second and will maintain this frequency level within ±1 cycle per second for the full load range and for the extremes of environmental conditions to be found in the field. We feel that such a system is practical on the basis of laboratory tests in which we have maintained the constant frequency as prescribed for loads from zero to full load and for temperature variations of −54 degrees centigrade to +71 degrees centigrade. On the basis of these assumptions we can utilize a relatively simple method of automatically closing the line circuit breaker at the correct time.

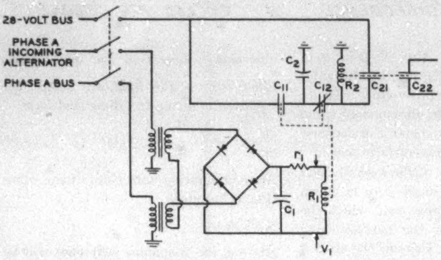
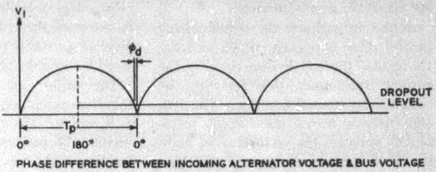


Figure 1 (left). Basic circuit

Figure 2 (below). Resulting wave form



The basic circuit is shown in Figure 1 of the discussion. With the paralleling switch closed the circuit senses the difference voltage between the same phase of the incoming alternator and the bus voltage. The resulting voltage, V_1, is rectified, filtered, and impressed across relay R_1. The input voltage polarities are such as to make V_1 zero for the in-phase condition. The filter passes the slowly varying voltage which is the beat frequency between the incoming alternator frequency and the bus frequency; it sharply attenuates the 800-cycle-per-second ripple frequency. The resulting wave form is shown in Figure 2 of the discussion.

Table I. Measured Data

ϕ_d, Degrees: +prior to 0°; −after 0°	Difference Frequency, Cycles per Second	Incoming Alternating Voltage	Bus Voltage	Record
+ 5	1.5	115	115	I
+10	0.2	115	115	II
−18	2.5	115	115	III
0	1	110	110	IV
−40	3.5	110	110	V
+15	0.2	110	110	VI
0	1	120	120	VII
−12	2.5	120	120	VIII

When V_1 is initially above the drop-out voltage level of relay R_1 it is energized, contact C_{11} is closed, C_{12} is open, and capacitor C_2 is charged to 28 volts. When V_1 decreases below the drop-out level of relay R_1 contact C_{11} opens, C_{12} closes, and capacitor C_2 discharges through relay R_2 causing it to pick up and lock in through its contact C_{21}. At the same time contact C_{22} closes the line circuit breaker paralleling the incoming alternator to the bus.

If V_1 happens to be below the drop-out level at the time the paralleling switch is initially closed, capacitor C_2 is not initially charged and relay R_2 cannot be energized. The circuitry thus provides assurance that relay R_2 will always be energized a constant number of degrees, ϕ_d, prior to the time of degrees phase difference between the incoming alternator voltage and the bus voltage. For the small angles involved, ϕ_d may be converted directly to time Δ_t where

$$\Delta_t = \phi_n T_p$$

when T_p = period of difference frequency.

If Δ_t is now made equal to the total delay time of both relay R_2 and the line circuit breaker for a difference frequency of 1 cycle per second, the incoming alternator will be paralleled at zero degrees phase difference. For difference frequencies above and below 1 cycle per second the phase difference between the incoming alternator and the bus voltages will be small. Table I of the discussion presents measured data of phase difference for typical difference frequencies and line voltages. On the basis of the initial assumption of a maximum difference frequency of 2 cycles per second, the data indicate the phase difference, ϕ_d, will probably always remain within 15 degrees of the in-phase condition.

In evaluating this automatic synchronizing method it is appreciated that close frequency control is vital. However, as mentioned previously, we believe this to be quite practical on the basis of our own experience. The suggested approach is the result of some very recent work during which time a variety of circuits were tried. One important fact connected with this approach is that only one such control box would be needed regardless of how many alternators were to be paralleled.

Russell W. Stineman (Boeing Airplane Company, Seattle, Wash.): The authors have presented a very neat and economical solution to the problem of developing a simpler paralleling procedure. It is likely that the same technique could be applied to other aircraft employing different load division circuits. The only question remaining is whether a problem really exists.

The postwar trend toward automaticity in weapons, especially aircraft, has been questioned in many quarters. Can this be another "gadget" whose weight detracts more from the vital performance characteristics of the aircraft than is justified by the derived benefits? The weight of this device alone is, of course, rather small, but the total weight of automatic devices on modern aircraft is definitely not negligible. It would appear that the answers to these questions will come only as more service experience is accumulated.

The need for automatic synchronizing is somewhat dependent on the alternator impedances. For example, one existing 60-kva alternator has larger per-unit synchronous and transient reactances than the 40-kva alternator. As a result, synchronizing currents on the 60-kva alternator are relatively small, and no exciter reversals or damage to drives have been experienced in our tests, even with random paralleling. For the 60-kva alternator, it appears that automatic synchronizing would be only a convenience.

Stability Limits for Third-Order Servomechanisms

T. J. HIGGINS
MEMBER AIEE

J. G. LEVINTHAL
NONMEMBER AIEE

IN VIEW of the important, numerous, and rapidly increasing uses of servomechanisms, analytic procedures that facilitate rapid design are of obvious technical value. The prime purpose of this paper is to set out the details of such a procedure. It has two major uses: First, in analysis it affords quick ascertainment of the roots of a specified cubic characteristic equation, thus enabling rapid determination of the stability characteristics of the corresponding third-order servo system; conversely, in synthesis it enables selection of the coefficients of a cubic characteristic equation to the end that a third-order servo system can be formulated with a desired degree and kind of stability.

Investigation or assignment of the stability properties of a servomechanism through study of the corresponding characteristic equation is of primary importance in the analysis, synthesis, or design of a servo system. Commonly this equation is algebraic with real coefficients. If, further, it is cubic, as very frequently occurs in practice, analysis through techniques now in use may involve tedious graphical or numerical procedures. These procedures can be obviated by use of the γ, α-, and ω-curves advanced in this paper.

In particular, these curves afford directly the three roots and the coefficients of a cubic characteristic equation such that the corresponding physical system has a desired kind and degree of stability and transient performance. Over the range for which the curves have been plotted, accuracy of root or coefficient determination is limited only by the scale to which the curves are plotted. Both increase of range and enlargement of scale can be effected easily by anyone so interested.

The essential content of this paper is to be summarized as follows. First, the theory underlying construction of the mentioned curves is set out. Second, construction of the curves is discussed in detail. Third, illustration of application of these curves to servomechanism design, analysis, and synthesis is advanced by numerical solution of several typical problems. Finally, sample curve sheets of the γ-, α-, and ω-curves are set out in the Appendix.

The corresponding curves for quadratic characteristic equations are given in several unpublished papers by J. F. Koenig, including one presented at the AIEE Great Lakes District Meeting, May 17–19, 1951. In these papers Koenig discusses the general theory of a procedure for delineating a certain desired performance of a servo system having an nth order characteristic equation and, illustratively, applies it to effect a detailed treatment of the general second-order characteristic equation. The considerable merit of this approach suggested the project of employing the mentioned general theory to effect, correspondingly, the following detailed treatment of the general cubic equation characterizing a third-order servo system.

Derivation of the γ-, α-, and ω Curves

ROOT GEOMETRY

Per well-known theory,[1,2] the nature of the stability and, in some measure, the transient performance of a linear physical system are determined by the nature of the roots of the characteristic equation of the system. Illustratively, consider the block diagram of the simple single-loop direct-feedback servomechanism of Figure 1. From this diagram we obtain by inspection

$$C(s)=G(s)E(s) \tag{1}$$

and

$$E(s)=R(s)-C(s) \tag{2}$$

Eliminating $E(s)$ between equations 1 and 2 yields

$$C(s)=\frac{G(s)}{1+G(s)}R(s) \tag{3}$$

The nature of the stability of this system is now obtainable by finding its response $c(t)$ to unit impulse $r(t)$; thus by taking $r(t)=\delta(t)$, whence $R(s)=1$, and inverting the resulting right-hand member of equation 3 through use, say, of partial fractions. Effecting this inversion requires determination of the roots of the characteristic equation

$$1+G(s)=0 \tag{4}$$

But as, in turn, the real parts of these roots furnish the decrement factors and the imaginary parts the natural angular frequencies of the exponentially-damped sinusoidal terms resulting from the partial fraction inversion, we have that the stability of the system is thus directly determined by the nature of the roots of the characteristic equation 4.

In similar fashion, it follows from equation 3 that the nature of the transient performance to an applied input $r(t)$ is determined by the roots of the characteristic equation 4 and (thus accounting for the preceding phrase, "in some measure") also the poles and zeros of $R(s)$.

In most current servomechanisms work the characteristic equation 4 is an algebraic equation with real coefficients. Confining attention to the cubic equation, let the canonical (standard) form be taken as

$$s^3+a_1s^2+a_2s+a_3=0 \tag{5}$$

In this equation s is the independent variable and a_1, a_2, and a_3 are real coefficients.

Equation 5 has three roots. All three roots may be real, or there may be one real root and a pair of complex conjugate roots. In the latter instance, the real parts of the conjugate roots may be greater or less than the real root (the case of real parts equal to the real root is encompassed as the limiting case). These three possibilities are indicated in Figure 2, whereon the roots are plotted in the $s=\sigma+j\omega'$ plane. All roots are indicated as having negative real parts: for as such must be the case if the associated system is to be stable in operation, we limit our discussion accordingly.

Now at least one root is always real, hence lies on the axis of reals. When only one such root occurs it is this root which is designated by $-\gamma$, see Figures 2(B) and 2(C). When three real roots occur, that nearest the axis of imaginaries is designated by $-\gamma$, see Figure 2(A). The absolute value of the imaginary part of each complex conjugate root is assigned the value ω.

Paper 52-313, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Middle Eastern District Meeting, Toledo, Ohio, October 28–30, 1952. Manuscript submitted July 28, 1952; made available for printing September 29, 1952.

T. J. HIGGINS is professor of electrical engineering, University of Wisconsin, Madison, Wis. J. G. LEVINTHAL is with the California Research and Development Corporation, Livermore, Calif.

This paper is based upon a thesis supervised by Professor Higgins and submitted by Mr. Levinthal to the faculty of the University of Wisconsin in partial fulfillment of the requirements for the degree of Master of Science in electrical engineering in June 1951.

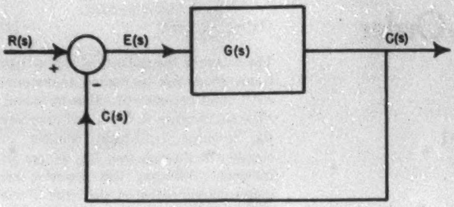


Figure 1. Basic notation and connections for a single-loop direct-feedback servomechanism

DETERMINATION OF THE γ- AND α-CURVES

Consider the effect of adding a real value α to each root exhibited in Figures 2(A), 2(B), and 2(C). If α is positive, each root in the s-plane is shifted to the right by a distance α. If α is negative, each root is shifted to the left by a distance α. Analytically, adding α to each root is equivalent to effecting the linear transformation

$$s=z-\alpha \tag{6}$$

Carrying out this transformation by substituting from equation 6 into equation 5 gives

$$(z-\alpha)^3+a_1(z-\alpha)^2+a_2(z-\alpha)+a_3=0 \tag{7}$$

Expanding and combining terms in equation 7 yields

$$z^3+(a_1-3\alpha)z^2+(3\alpha^2-2a_1\alpha+a_2)z+ (a_1\alpha^2-a_2\alpha+a_3-\alpha^3)=0 \tag{8}$$

If now in Figure 2(A) we take $\alpha=\gamma$, in Figure 2(B) we take $\alpha=\epsilon$, and in Figure 2(C) we take $\alpha=\gamma$, then equation 8 has possible roots as shown in Figure 3. In such case equation 8 has at least one root on the imaginary axis, and none in the right half-plane. Now such location of roots gives rise to special relations among the coefficients of equation 8. Thus, inasmuch as equation 8 is of the form

$$z^3+b_1z^2+b_2z+b_3=0 \tag{9}$$

it is evident that if a root is to be located at the origin, as in Figures 3(A) and 3(C), it must be that

$$b_3=0 \tag{10}$$

Again, for roots located as in Figure 3(B), it follows from well-known theory that the second and third Hurwitz determinants of equation 9 must vanish. Hence, for such root location, we must have also that

$$\begin{vmatrix} b_2 & b_3 \\ 1 & b_1 \end{vmatrix}=0 \tag{11}$$

Now, from comparison of equations 8 and 9 we have

$$b_3=a_1\alpha^2-a_2\alpha+a_3-\alpha^3 \tag{12}$$

Substituting equation 12 in equation 10 and solving for a_3 yields

$$a_3=\alpha^3+a_2\alpha-a_1\alpha^2 \tag{13}$$

This equation determines the value of α such that under the transformation $s=z-\alpha$ wherein $\alpha=\gamma$, the real root of the transformed equation 8 will fall at the origin. Accordingly, we have that the value of the root $s=-\gamma$ of the original equation 5 is to be obtained by substituting γ for α in equation 13, thus

$$a_3=\gamma^3+a_2\gamma-a_1\gamma^2 \tag{14}$$

and solving equation 14 for γ.

Equation 14, if interpreted as an expression for a_3 as a function of a_2, wherein a_1 and γ are parameters of constant value, is the equation of a double-parameter family of straight lines. A collection of these curves, constructed as will be discussed in detail, comprises the γ-curves, typified by curve sheets A-3, A′-3, and A′-4 in Figures 7, 8, and 9. These curves readily afford the solution of equation 14 for γ.

From equations 8 and 9 we have

$$\begin{aligned} b_1&=a_1-3\alpha \\ b_2&=3\alpha^2-2a_1\alpha+a_2 \\ b_3&=a_1\alpha^2-a_2\alpha+a_3-\alpha^3 \end{aligned} \tag{15}$$

Substituting these values in the determinant of equation 11 yields

$$\begin{vmatrix} (3\alpha^2-2a_1\alpha+a_2) & (a_1\alpha^2-a_2\alpha-\alpha^3) \\ 1 & (a_1-3\alpha) \end{vmatrix}=0 \tag{16}$$

On expanding the determinant in equation 16 we obtain

$$(3\alpha^2-2a_1\alpha+a_2)(a_1-3\alpha)- (a_1\alpha^2-a_2\alpha+a_3-\alpha^3)=0 \tag{17}$$

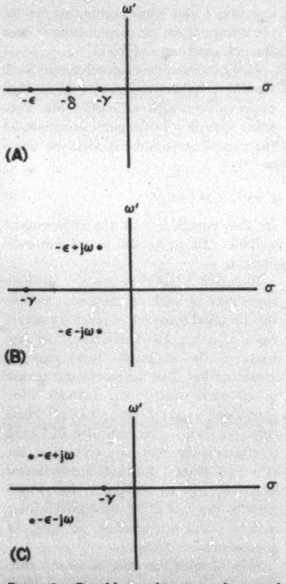


Figure 2. Possible combinations of roots of the canonical equation 5, $s^3+a_1s^2+a_2s+a_3=0$

A. Equation 5 has three real roots

B. Equation 5 has one real root and a pair of complex conjugate roots to the right of the real root

C. Equation 5 has one real root and a pair of complex conjugate roots to the left of the real root

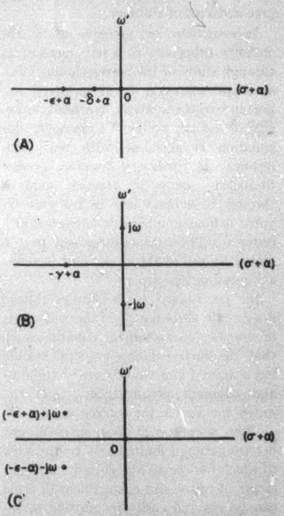


Figure 3. Possible combinations of roots in the $z=s+\alpha=(\sigma+\alpha)+j\omega'$ plane of the transformed canonical equation $z^3+z^2(a_1-3\alpha)+z(3\alpha^2-2a_1\alpha+a_2)+(a_1\alpha^2-a_2\alpha+a_3-\alpha^3)=0$

A. Equation 5 has three real roots

B. Equation 5 has one real root and a pair of complex conjugate roots to the right of the real root

C. Equation 5 has one real root and a pair of complex conjugate roots to the left of the real root

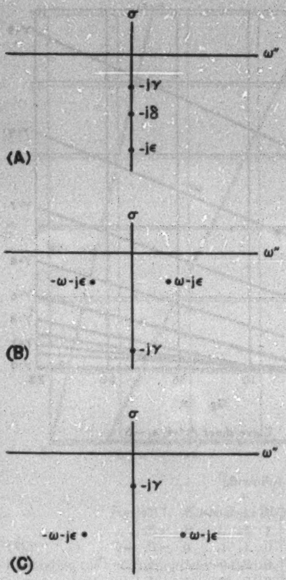


Figure 4. Possible combinations of roots in the $q=js=-\omega'+j\sigma=\omega''+j\sigma$ plane of the roots of the transformed canonical equation $(-jq)^3+a_1(jq)^2+a_2(jq)+a_3=0$

A. Equation 5 has three real roots
B. Equation 5 has one real root and a pair of complex conjugate roots to the right of the real root
C. Equation 5 has one real root and a pair of complex conjugate roots to the left of the real root

Expanding this equation yields

$$3a_1\alpha^2-2a_1^2\alpha+a_1a_2-9\alpha^3+6a_1\alpha^2-3a_2\alpha-a_1\alpha^2+a_2\alpha-a_3+\alpha^3=0 \quad (18)$$

Solving equation 18 for a_3 results in

$$a_3=(a_1-2\alpha)a_2-8\alpha^3+8a_1\alpha^2-2a_1^2\alpha \quad (19)$$

This equation determines the value of α such that under the transformation $s=z-\alpha$, the complex roots of the transformed equation 8 will fall on the imaginary axis. Accordingly, we have that the value $-\epsilon$ of the real parts of the complex roots of the original equation 5 is to be obtained by solving equation 19 and taking $\alpha=\epsilon$.

Equation 19, if interpreted as an expression for a_3 as a function of a_2 wherein a_1 and α are constant parameters, defines a double-parameter family of straight lines. A collection of curves, constructed as will be discussed in detail later, comprises the α-curves, typified by curve sheets *B*-3, *B'*-3, and *B'*-4 in Figures 10, 11, and 12. These curves readily afford the solution of equation 19 for α.

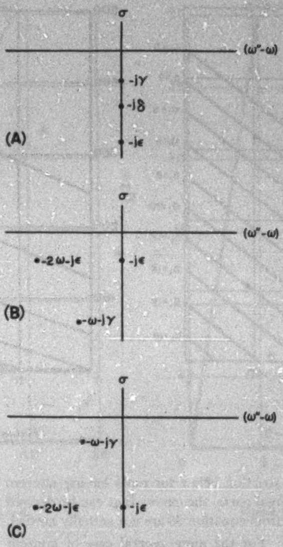


Figure 5. Possible combinations of roots in the $p=q-\omega=(\omega''-\omega)+j\sigma$ plane of the roots of the transformed canonical equation $p^3+p^2(3\omega+ja_1)+p[(3\omega^2-a_2)+2a_1\omega]+[(\omega^3-a_2\omega)+j(a_1\omega^2-a_3)]=0$

A. Equation 5 has three real roots
B. Equation 5 has one real root and a pair of complex conjugate roots to the right of the real root
C. Equation 5 has one real root and a pair of complex conjugate roots to the left of the real root

DETERMINATION OF THE ω-CURVES

The γ- and α-curves can be interpreted as relating the real parts of the roots of the characteristic equation to the coefficients of this equation. We turn now to determination of a similar set of curves for the imaginary parts of the roots.

Consider the *s*-plane of Figure 2 as rotated counterclockwise by 90 degrees. This rotation is analytically equivalent to performing the transformation.

$$s=-jq \quad (20)$$

The *q*-plane resulting from such transformation is shown in Figure 4. In turn, shifting points in the *q*-plane horizontally to the left by the indicated positive distance ω places all roots on or to the left of the vertical axis of the *p*-plane, as shown in Figure 5. Analytically, this is equivalent to performing the transformation.

$$q=p+\omega \quad (21)$$

From equations 20 and 21 we obtain

$$s=-j(p+\omega) \quad (22)$$

enabling direct transformation from the *s*-plane to the *p*-plane.

Figure 5 indicates the nature of the roots on the *p*-plane. Either one or three roots are on the imaginary axis, and in no case is a root to the right of this axis. Further, should the real part of a complex root in the *s*-plane be zero, the transformed roots lie on the real axis of the *p*-plane.

Substituting equation 22 in equation 5 gives

$$[-j(p+\omega)]^3+a_1[-j(p+\omega)]^2+a_2[-j(p+\omega)]+a_3=0 \quad (23)$$

Expanding, multiplying by j, and collecting terms yields

$$p^3+p^2[3\omega+ja_1]+p[(3\omega^2-a_2)+j2a_1\omega]+[(\omega^3-a_2\omega)+j(a_1\omega^2-a_3)]=0 \quad (24)$$

This equation is of the form

$$p^3+p^2(A_1+jB_1)+p(A_2+jB_2)+(A_3+jB_3)=0 \quad (25)$$

Now location of roots such as in Figure 5 gives rise to certain relations among the coefficients. These relationships can be obtained from the Bilharz-Frank criterion[2] ensuring that all roots of an equation with complex coefficients have negative real parts. Thus, for the case of

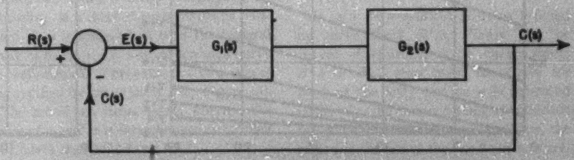


Figure 6. Block diagram of a single-loop direct-feedback servomechanism

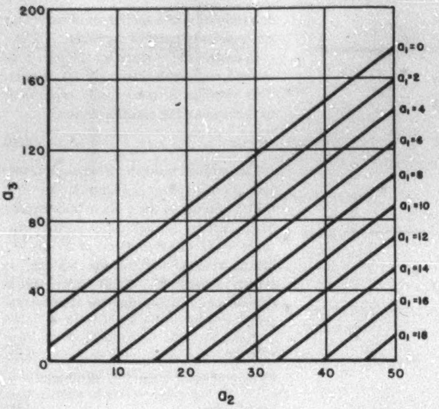


Figure 7. Curve sheet A-3, $\gamma=3$

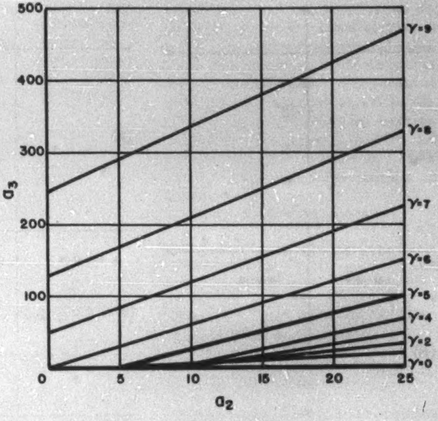


Figure 9. Curve sheet A′-4, $a_1=6$

roots with no real parts it is necessary that

$A_3=0$ and $B_3=0$ (26)

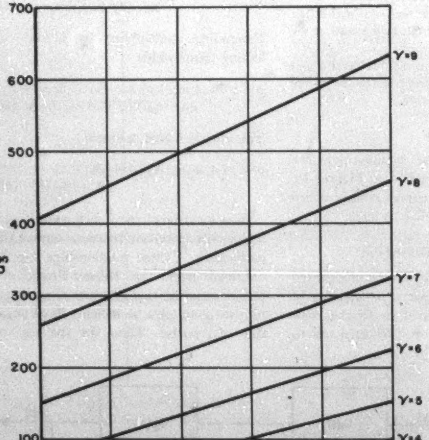
Since roots that satisfy this restriction also will satisfy the more general restriction below for roots having nonzero real parts, the curves that can be derived from equation 26 are not actually needed.

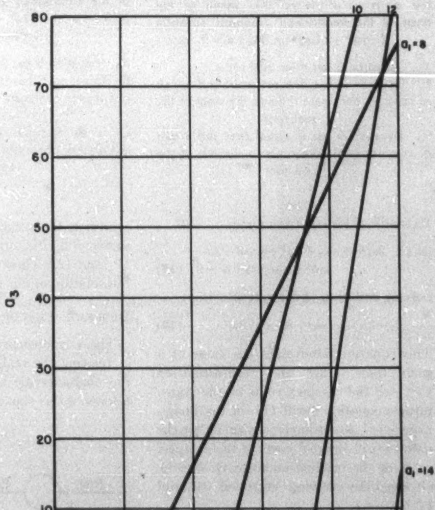
For the more general case of nonzero real parts, application of the Bilharz-Frank criterion to equation 25 yields the following

$$\begin{vmatrix} A_1 & A_3 & 0 & -B_2 & 0 \\ 1 & A_2 & 0 & -B_1 & -B_3 \\ 0 & A_1 & A_3 & 0 & -B_2 \\ 0 & B_2 & 0 & A_1 & A_3 \\ 0 & B_1 & B_3 & 1 & A_2 \end{vmatrix} = 0 \qquad (27)$$

Figure 8 (below). Curve sheet A′-3, $a_1=4$

Figure 10 (right). Curve sheet B-3, $\alpha=3$





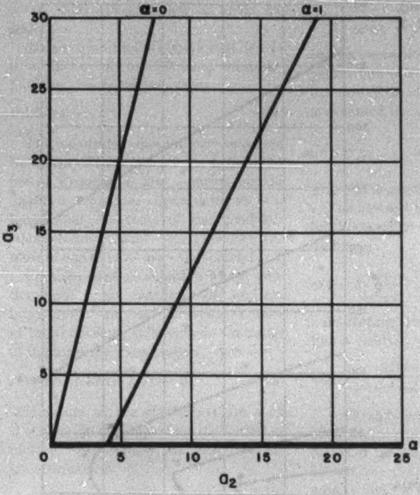


Figure 11. Curve sheet B′-3, $a_1=4$

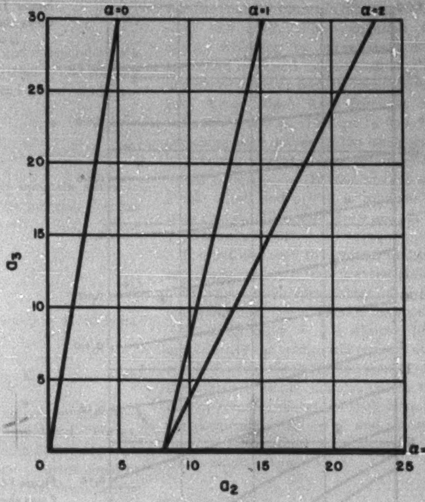


Figure 12. Curve sheet B′-4, $a_1=6$

Expanding the determinant and collecting terms results in

$$[A_1A_2-A_3][A_1A_2A_3+A_1B_2B_3-A_3^2]+[B_2-A_1B_1][A_3^2B_1-B_2^2B_3-A_2A_3B_2-A_1A_3B_3]+A_1B_3[A_1A_3B_1-A_3B_2-A_1^2B_3]=0 \quad (28)$$

Substituting the coefficients of equation 24 and collecting terms yields

$$27a_3^2+(4a_1^3-18a_1a_2)a_3+4a_2^3-(36\omega^2+a_1^2)a_2^2+a_2(96\omega^4+24a_1^2\omega^2)-64\omega^6-32a_1^2\omega^4-4a_1^4\omega^2=0 \quad (29)$$

Solving for a_3 gives

$$a_3=(1/27)[(9a_1a_2-2a_1^3)+\{(1/2)(2a_1^3-9a_1a_2)^2+27[-4a_2^3+(36\omega^2+a_1^2)-(96\omega^4+24a_1^2\omega^2)a_2+64\omega^6+32a_1^2\omega^4+4a_1^4]\}^{1/2}] \quad (30)$$

Equation 30, interpreted as an expression for a_3 as a function of a_2 wherein a_1 and ω are parameters of constant value, is the equation of a family of curves of rather simple form. Although the right-hand side of equation 32 is a complicated expression, points for a family of curves, constructed in detail as discussed later, can be obtained without too much labor by use of International Business Machines computing equipment. A collection of these curves comprises the ω-curves, typified by curve sheets C-3, C'-3, and C'-4 in Figures 13, 14, and 15. These curves readily afford solution of equation 30 for ω, therefore providing the imaginary parts of the complex roots of equation 5.

Construction of the γ-, α-, and ω-Curves

If γ in equation 14 is held constant, a_1 assigned in turn the integral values 0, 2, . . . , 18, 20, and curves of a_3 versus a_2 plotted (over the first quadrant of the a_3, a_2 plane) on common axes, a curve sheet containing a family of eleven straight lines results. If a sheet of such curves is plotted as γ is assigned in turn the integral values 0, 1, . . . , 9, a set of ten such curve sheets ensues. This collection of γ-curves is typified by curve sheet A-3. For determination of γ from equation 14, it is somewhat more convenient to use the set of curve sheets to be obtained by replotting those curves having a common fixed value of a_1 on individual curve sheets, whereon γ is the parameter. This collection of γ-curves is typified by curve sheets A'-3 and A'-4.

If, similarly, α in equation 19 is held constant, a_1 assigned in turn appropriately chosen integral values, and curves of a_3 versus a_2 plotted on common axes a curve sheet containing a family of straight lines results. If a sheet of such curves be plotted as α is assigned in turn the integral values 0, 1, . . . , 9, a set of ten such curve sheets ensues. This collection of α-curves is typified by curve sheet B-3, Figure 10. The set of ten curve sheets whereon the role of a_1 and α is interchanged is typified by curve sheets B'-3 and B'-4.

A similar set of ten curve sheets, plotted from equation 30, each for an integral value of $\omega=0$, 1, . . . , 9, and the corresponding interchanged set yields the collection of ω-curves typified by curve sheets C-3, C'-3, and C'-4.

The range of values encompassed on these curve sheets is such that by conjunction of interpolation among these curves and by simple linear transformation of the variable s as indicated in the following illustrative problem, most problems arising in practice can be solved directly therefrom. Obviously, however, these curves can be replotted to any desired scale or extended over any desired range by anyone so interested. In this connection it is to be remarked that although sheer bulk precludes advancing in this paper the total of sixty curve sheets encompassing all of the γ-, α-, and ω-curves, the thesis containing these curves can be borrowed and the latter reproduced *in toto* by an interested party.

Applications of the Curves

STABILITY LIMITS

It is to be remarked that, as typified by curve sheet B'-3, Figure 11, the curves of constant α on curve sheets B'-1 to B'-7 start with the curve for $\alpha=0$ at the left; and as α is increased positively, the corresponding curves shift to the right of that for $\alpha=0$. Thus, the region to the left of the $\alpha=0$ curve corresponds to negative values of α. How-

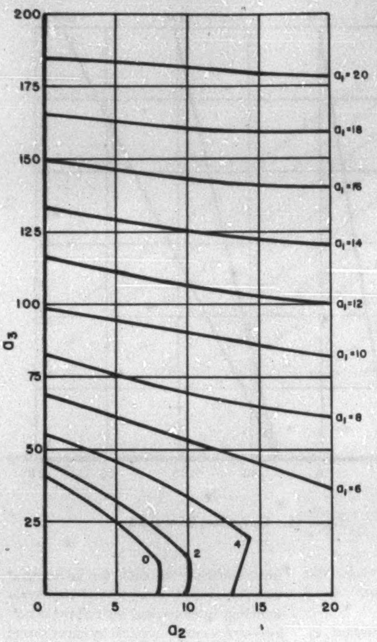


Figure 13 (left). Curve sheet C-3, $\omega=3$

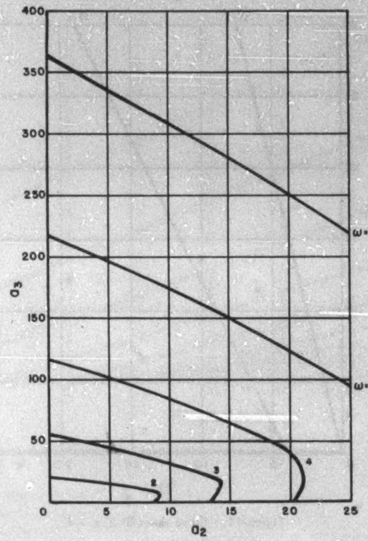


Figure 14 (right). Curve sheet C′-3, $a_1=4$

ever, inasmuch as a negative value of α corresponds to a real positive root of the characteristic equation 5, root location in this region can occur only for a system that is unstable. The $\alpha=0$ curve thus determines the limit of stability of a system. Setting $\alpha=0$ in equation 19 and solving for the ratio of a_3 to a_2 gives

$$a_3/a_2=a_1 \quad (31)$$

For a system to be stable, then, it is necessary that

$$a_3/a_2<a_1 \quad (32)$$

Confirmatively, this condition is identical with the stability requirement stemming from application of Routh's criterion to the cubic characteristic equation.

DETERMINING THE ROOTS OF CUBIC EQUATIONS

Consider first an equation having all real roots

$$s^3+6s^2+11s+6=0 \quad (33)$$

In this example $a_1=6$, $a_2=11$, and $a_3=6$. Knowing one root must be real, curve sheet A'-4, for which $a_1=6$, is first utilized. From this sheet it is found that $\gamma=1$, whence -1 is the real root. From curve sheet C'-4 for which $a_1=6$ it is found that

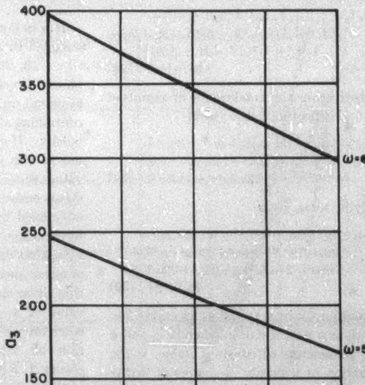


Figure 15 (right). Curve sheet C′-4, $a_1=6$

$\omega=0$. This indicates that all roots are real.

Since -1 is a root of equation 33, $(s+1)$ is a factor of its left-hand member. Dividing this factor out yields

$$s^2+5s+6=0 \qquad (34)$$

This quadratic equation can be solved by means of the quadratic formula, or by use of Koenig's α- and ω-curves for the quadratic equation. Upon doing so, it is found that the roots of the quadratic equation are -2 and -3. Therefore, the roots of equation 33 are -1, -2, and -3.

Consider next an equation whose solution both yields complex roots and exhibits how to proceed when one or more of the coefficients falls outside the range of the curve sheets as plotted; thus

$$s^3+40s^2+1{,}400s+20{,}000=0 \qquad (35)$$

The values of the coefficients are such that they do not appear on the curves as plotted. Accordingly, we introduce the transformation

$$s=10u \qquad (36)$$

Equation 35 then goes into

$$1{,}000u^3+4{,}000u^2+14{,}000u+20{,}000=0 \qquad (37)$$

Dividing by 1,000 yields

$$u^3+4u^2+14u+20=0 \qquad (38)$$

which falls within the range of the curve sheets.

As one root must be real, curve sheet A'-3 for which $a_1=2$ is first utilized. From this sheet it is found that $\gamma=2$, whence one real root $=-2$. From curve sheet C'-3 for which $a_1=2$, it is found that $\omega=3$. Hence, there are complex roots. To ascertain their real parts we utilize curve sheet B'-3, for which $a_1=2$. It is found therefrom that $\alpha=1$, whence the real part of the complex roots is -1. Accordingly, the roots of equation 38 are $u=-2$, $u=-1+j3$, and $u=-1-j3$. Substituting these values in equation 36 gives the roots of the original equation 35. Thus, $s=-20$, $s=-10+j30$, and $s=-10-j30$.

These two examples illustrate all that is involved in solving cubic characteristic equations by use of the γ-, α-, and ω-curves.

SYNTHESIS OF A SERVOMECHANISM

Figure 6 comprises the block diagram of a simple servo system. The two blocks might well indicate a corrective lead network and an idealized motor.

A typical expression for a resistance-capacitance lead network is given by the transfer function

$$G_1(s)=\frac{bs+c}{bs+1} \qquad (39)$$

In this expression b and c depend on the parameters of the system. The transfer function of an idealized motor can be approximated by

$$G_2(s)=\frac{A}{s^2} \qquad (40)$$

where A depends only upon the motor.

The forward transfer function $G(s)$ can be expressed as $G(s)=G_1(s)G_2(s)$. Hence we have

$$G(s)=A\frac{(bs+c)}{bs^3+s^2} \qquad (41)$$

Substituting from equation 41 in equation 5 yields

$$C(s)=\frac{A(bs+c)}{bs^3+s^2+Abs+Ac}R(s) \qquad (42)$$

With equation 42 in hand several types of problems can be formulated. For example, if the coefficients A, b, and c were known, the γ-, α-, and ω-curves could be used to find the roots of the characteristic equation

$$bs^3+s^2+Abs+Ac=0 \qquad (43)$$

as in the just completed illustrative examples.

More general interest attaches to the following synthesis problem. Assume that a reasonable value of b is 0.25, that for a step input the error must be less than 5 per cent within a period of 3 seconds, and that the response shall have a sinusoidal component of angular velocity $\omega=2$ radians per second. Now, by well-known theory: the magnitude of the real part $-\alpha$ of a root of equation 5 is equal to the reciprocal of a time constant of the system; and, further, in a time equal to three times this time constant the exponential harmonic term resulting from a pair of conjugate roots will have its magnitude reduced to at least as much as is desired. As a first effort, assume that the thus determined maximum $\alpha=1$ and the $\omega=2$ values occur for the pair of conjugate roots.

Substituting the known value of b in equation 42 yields

$$C(s)=\frac{A(0.25s+c)}{0.25s^3+s^2+0.25As+Ac}R(s) \qquad (44)$$

Multiplying the numerator and denominator by 4 yields

$$C(s)=\frac{A(s+c)}{s^3+4s^2+As+Ac}R(s) \qquad (45)$$

The characteristic equation is given by

$$s^3+4s^2+As+Ac=0 \qquad (46)$$

For the complex roots of equation 46 try $\alpha=1$ and $\omega=2$. From equation 46, $a_1=4$. Referring to curve sheet B'-3 and curve $\alpha=1$ and to curve sheet C'-3 and curve $\omega=2$, we seek the point on the two curves which have common coordinates a_2 and a_3. This can be done by sketching both curves to the same scale and seeking a point of intersection, or by visual comparison. In the present example, $a_2=9$ and $a_3=10$ is common to both curves. Therefore $A=9$ and $AC=10$, whence $c=1.11$.

It is yet necessary to ascertain that the remaining real root satisfies the stated conditions. From curve sheet A'-3 we determine that $\gamma=2$, whence the root is -2. Accordingly, if A is adjusted to 9 and c to 1.11 the system will perform as is required. In general, the dominant α and the stated ω would not conjoin in the pair of complex roots. In such instance the dominant α will be γ, and the real parts of the complex roots having $\omega=2$ will be smaller than $-\gamma$.

Summary

The illustrative problems sufficiently illustrate the uses of the γ-, α-, and ω-curves. In general, both accuracy of solution and speed of calculation are excellent. The range of coefficients over which curves are plotted is, of course, limited. This is not too important a limitation for, as illustrated in one of the numerical examples, suitable substitutions can be found which transform the coefficients to values within the range of the curves. Again, the curves are plotted only for positive coefficients. Although this is the case of practical interest (a necessary restriction for a stable system is that all coefficients be of the same sign), it is possible to use the general equations for the curves to extend them to encompass negative coefficients, should such prove of interest.

Finally it is to be noted that γ-, α-, and ω-curves have been developed to date only for solution of second and third order servo systems. However, although the computational work involved in solving the equations pertinent to a higher order system is considerable, it is nonetheless a routine process that can be carried out and the corresponding curves obtained, if such are desired. In particular, effectance of such a set of curves for the quartic equation is not only feasible, but the resulting curves ought prove most useful in practice, particularly in connection with aircraft stability problems whereof a quartic characteristic equation is of major occurrence.[4]

Appendix

Levinthal's[5] thesis, the essential content of which comprises this paper, contains the sixty curve sheets described in the section, "Construction of the γ-, α-, and ω-Curves." These curve sheets are typified by the nine curve sheets *A*-3, *A'*-3, *A'*-4, *B*-3, *B'*-3, *B'*-4, *C*-3, *C'*-3, and *C'*-4 of Figures 7–15. However, unlike these latter curve sheets, those in the thesis are plotted on centimeter-squares graph paper, whence they can be read with considerable accuracy.

In course of review of this paper, one reviewer advanced the desirability of mentioning the charts given by Evans[6] and by Liu.[7] In addition to these charts, which are of somewhat different usefulness than those advanced in this paper, the authors would also remark the availability of yet other graphical and numerical charts to be found in papers by Oppelt,[8] Stefaniak,[9] and Jones.[10] The material encompassed in references 6 to 10 is of values in the analysis of third-order servomechanisms in the various fashions manifest in the titles of these papers.

References

1. Principles of Servomechanisms (book), **G. S. Brown, D. P. Campbell.** John Wiley and Sons, Inc., New York, N. Y., 1948, pages 170–76.

2. Servomechanisms and Regulating System Design (book), **H. Chestnut, R. W. Mayer.** John Wiley and Sons, Inc., New York, N. Y., volume 1, 1951, pages 124–30.

3. The Geometry of the Zeros of a Polynomial in a Complex Variable book), **M. Marden.** American Mathematical Society, New York, N. Y., 1949, page 140.

4. The Principles of the Control and Stability of Aircraft (book), **W. J. Duncan.** Cambridge University Press, Cambridge, England, 1952, page 116.

5. Stability Limits for Third-Order Servomechanisms, **J. G. Levinthal.** *M.S. Thesis*, University of Wisconsin (Madison, Wis.), June 1951.

6. Solution of the Cubic Equation and the Cubic Charts, **L. W. Evans.** *Monograph*, Massachusetts Institute of Technology (Cambridge, Mass.), 1941, pages 50–53.

7. Servomechanisms: Charts for Verifying Their Stability and for Finding the Roots of Their Third and Fourth Degree Characteristic Equations, **Y. J. Liu.** *Monograph*, Massachusetts Institute of Technology (Cambridge, Mass.), 1941, pages 55–74.

8. Zum Dämpfungsgrad der Regelungsdifferentialgleichung Dritter Ordnung, **W. Oppelt.** *Archiv für Elektrotechnik* (Berlin, Germany), volume 37, 1943, pages 357–60.

9. Ein Graphisches Verfahren Zur Bestimmung der Zeitkonstanten und der Schwingungsdauer eines Linearen Systemes Dritter Ordnung, **H. St. Stefaniak.** *Ingenieur-Archiv* (Berlin, Germany), volume 18, 1950, pages 221–32.

10. Solution of the Cubic Equation by a Simplified Procedure, **E. E. Jones.** *Product Engineering* (New York, N. Y.), volume 23, number 9, September 1952, pages 183–89.

Discussion

J. F. Koenig (National Bureau of Standards, Washington, D. C.): The authors have applied a general method to the specific case of third-order servomechanisms. They have not referred to another paper[1] which also applies the root trajectory method to the specific case of third-order servomechanisms. In this paper[1] two root trajectory diagrams, giving all roots of the characteristic equation, were plotted in terms of two physical parameters, not coefficients of the characteristic equation. One parameter is K and the other r. The coefficient $a_2 = K$ and $a_3 = K/(1+r)$. $a_1 = 1$. The diagrams can be replotted easily, in terms of a_2 and a_3, giving all roots of the general third-order system. a_1 always can be made equal to unity by use of the simple method which is given on page 136 *et seq.* of reference 2 of the paper.

The authors could have plotted their diagrams in terms of two coefficients, for example, a_2 and a_3. Then the number of curve sheets would be reduced from 60 to 2: one γ, and the superposition of one ω and one α. The total effect of all parameter variation would be expressed in terms of two coefficients. There is considerable value in plotting the diagrams in terms of physical parameters, rather than coefficients, since the effect of parameter variation on the root trajectories becomes evident.

In equation 27 the authors equated the last Bilharz-Frank determinant Δ_n to zero. They are mistaken in stating that this is an application of the Bilharz-Frank criterion. The equating of Δ_n to zero is the application of a theorem developed by the writer. It states: "If any *n*th degree complex polynomial equation (such as the authors' equation 25) has at least one pure imaginary root, then the last Bilharz-Frank determinant Δ_n is zero." The authors are mistaken when they imply that the converse of this theorem is always true. Fortunately, however, the special cases for which the converse is not true are believed to be of little practical significance.

Elsewhere the writer has superposed ω curves and α curves. The authors could have superposed some of their diagrams, for example, Figures 11 and 14, in order to have a clearer picture of the root trajectories as a function of coefficient variation.

The authors have not mentioned that families of constant damping ratio, that is, $\tan\beta = |\alpha/\omega|$, can be plotted. The writer plotted curves of constant β in an unpublished paper, for the quadratic system, and another paper[1] for the cubic system. The values of the complex roots are found from the intersections of the α curves with the β curves.

In the paragraph immediately following equation 14, the authors are mistaken in their belief that any one of the curve sheets typified by *A'*-3 and *A'*-4 gives only $-\gamma$, the least negative real root. Actually any one will give all the real roots. For example, the authors use *A'*-4 for equation 33 to obtain a real root $-\gamma = -1$. However, the three curves, $\gamma = 1, 2, 3$, pass through the point $a_2 = 11$, $a_3 = 6$ of *A'*-4. Therefore *A'*-4 gives all three real roots, provided that a careful plot is made. This principle has been developed and applied[1] by the writer: All the real roots of a real or complex polynomial equation of the *n*th degree in the complex variable s can be obtained by substituting $s = z - \gamma$, equating to zero the constant term of the z-equation, and plotting the resulting equation.

In equations 11, 16, and 19 the authors have equated the next to last Hurwitz determinant H_{n-1} to zero. In the sentence immediately following equation 19 the authors are mistaken when they state that equation 19 determines the value of α such that under the transformation $s = z - \alpha$, the complex roots of the transformed equation 8 will fall on the imaginary axis. Actually there are special cases for which equation 19 (or $H_{n-1} = 0$ for any *n*th degree real polynomial equation) is satisfied although equation 8 (or any *n*th degree real polynomial equation) does not have a pure imaginary pair of roots.

Reference

1. Design of Damper-Stabilized Instrument Servomechanisms, **J. F. Koenig.** *Proceedings*, National Electronics Conference (Chicago, Ill.), October 22–24, volume 7, 1951.

T. J. Higgins and **J. G. Levinthal:** We did not mention reference 1 of Mr. Koenig's discussion as our paper essentially encompasses the content of Mr. Levinthal's thesis (reference 5), which was finished late in May or early in June of 1951. However, as this reference does contain certain curves which are particularly useful for the design of third-order damper-stabilized servomechanisms, the authors are pleased that it is now directed to the attention of those interested in this topic through Mr. Koenig's citing of it.

As Mr. Koenig remarks, the curves of this paper are plotted as functions of the specialized parameters K and r occurring in damper-stabilized theory whence direct use of the curves is restricted accordingly. Of course, as Mr. Koenig notes, these curves can be reworked to yield a new set of curves plotted as functions of a_2 and a_3, which give the roots of a restricted class of third-order cubics (namely, those for which $a_1 = 1$). From these curves can then be obtained by use of the "simple method" given on page 136 of reference 2 of our paper the roots of a cubic whereof $a_1 \neq 1$. Naturally, this particular scheme has been much utilized. For the mentioned method of reducing the number of arbitrary coefficients from three to two through reducing one of a_1, a_2, or a_3 to unity by an appropriate change of the independent variable of the associated differential equation of which the cubic is the characteristic equation, or by change of the variable in the cubic itself, is an old one. Thus, to cite an early instance of use in technical work, precisely such a change was utilized by Wischnegradski[1] in 1877 to obtain a stability chart apposite to his investigation of a certain regulator problem characterized by a cubic equation. And to site a few instances in recent literature, precisely such procedure is utilized in obtaining the charts given by Evans,[2] Liu,[3] Grammel,[4] and Stefaniak.[5] Now of these charts, some are plotted as functions of two parameters particularly apposite to the corresponding subject under investigation, and others are plotted in terms of two of the coefficients—whence both of the possibilities suggested by Mr. Koenig are encompassed.

Although inasmuch as thorough knowledge

of the just-mentioned charts the authors were familiar with the fact that conjunction of use of the simple method with curves plotted in terms of a_1 and a_2 for $a_3 = 1$, or in terms of a_2 and a_3 for $a_1 = 1$, would have resulted in terms of two curve sheets rather than 60, our actual aim was otherwise. For aside from the fact that nothing new would have resulted therefrom (the plot in terms of a_1 and a_2 is already effected in Stefaniak's paper[5]), our prime purpose was to effect a set of curves in three parameters a_1, a_2, and a_3, and thus have a set of curves which can be used for a cubic without the necessity of an associated computation, as required by previous charts.

Again, whether or not there is "considerable value in plotting the diagrams in terms of physical parameters, rather than coefficients" depends on the use to which the charts are to be put. If prime attention centers on such parameters—as, for example, is the case for K and r of reference 1 of Mr. Koenig's discussion or, again, as is the case when the damping constant and frequency constant δ and w_r are of prime interest (as is the case manifest in chart 2 of reference 6 of the paper)—then, of course, these parameters should be chosen as the variables of the charts. But if, as is the case in our paper, prime attention focuses on the coefficients of the characteristic equation itself, then the "considerable" value lies in utilizing the coefficients as the variables of the charts.

The mentioned theorem can be utilized to set up equation 27 of the paper. It happens however, that such procedure is not unique. Thus, the problem of ascertaining conditions evidencing the nature of the roots of equation 25 was investigated long ago. In particular, in 1918 Bateman[6] set out both a general procedure and specific details for the quadratic and cubic. This work is easiest epitomized here by advance of Bateman's pertinent result for the quadratic. Let the roots be $-(x_1 + jy_1)$ and $-(x_2 + jy_2)$. Then through the procedure mentioned in the following paragraph Bateman obtains two expressions for a certain H-function: in terms of the roots, $H = x_1x_2[(x_1+x_2)^2+(y_1-y_2)^2]$; and in terms of the real and imaginary parts of the coefficients of the quadratic (utilizing the notation of our paper), $H = A_1B_1B_2 - B_2^2 + A_1^2A_2$. Now from the first expression for the H-function we have at once that if either one or both roots have zero real parts $(x_1, x_2 = 0)$, then H necessarily vanishes. And hence a necessary condition (and in this case a sufficient condition) satisfied by the coefficients when one or both roots have zero real parts is given by equating the second form of the H-function to zero.

Now Bateman's formulation for equations with complex coefficients is effected by an obvious analogous extension of the procedure Routh[7] used in investigating equations with real coefficients. This immediately suggests the following. For real-coefficient equations numerous investigators have manifested the common bases of the Routhean functions and the Hurwitzian determinants to the end that conditions formulated in terms of Routhean functions can immediately be paralleled in terms of Hurwitzian determinants and vice versa, and the same holds for complex-coefficient equations whereof the Routhean functions find their counterparts in the Bilharz-Frank determinants. In this thought, a glance at the Routhean H-function, $H = A_1B_1B_2 - B_2^2 + A_1^2A_2$, reveals that it is the expansion of the Bilharz-Frank determinant Δ_2, the highest-order determinant of the quadratic; whence we have at once a necessary condition: that if a quadratic with complex coefficients is to have one or both roots with zero real parts, then necessarily its highest-order Bilharz-Frank determinant vanishes. In the same fashion one can ascertain similarly—or even easier, conjecture and prove by induction—that if equation 25 is to have one or more roots with zero real parts, then necessarily the highest order Bilharz-Frank determinant vanishes, thus yielding equation 27.

Probably another way of arriving at the same result is to proceed in much the same fashion Sponder[8] does in proving that if a real-coefficient equation has one or more roots with zero real parts, then necessarily the highest-order Hurwitzian determinant vanishes. And, in fact, in knowledge of Sponder's work and of the fact that if all the coefficients of a complex-coefficient equation are real the corresponding Bilharz-Frank determinants reduce to the Hurwitzian determinants, one might immediately conjecture—in exact analogy of Sponder's result—that if a complex-coefficient equation is to have one or more roots with zero real parts, then necessarily the highest-order Bilharz-Frank determinant vanishes, proceed to test this by trial for the first-, second-, and third-order equations, and thus by induction for the nth-order equation.

Probably the authors' use of the phrase "an application of the Bilharz-Frank criterion to equation 25" was inadvisable; for as happened, its literal interpretation is of a more restricted sense then intended. No doubt use of a more appropriate phrase such as "an application of Bilharz-Frank analysis to equation 25" would have obviated Mr. Koenig's remarks on our procedure.

As Mr. Koenig remarks, the authors could have superimposed corresponding sets of α and ω curves, such as those of Figures 11 and 14 of the paper, through plotting on common axes. Again, however, they purposely avoided so doing. It is more convenient to maintain separate diagrams so as to avoid the necessity of working with a crowded confusion of two sets of curves in those instances when only one of a set is desired, and to obtain superposition when both sets are to be used jointly through superimposing the thin or translucent graphs whereon the curves are plotted.

The authors did not mention the possibility of plotting families of constant damping ratio, tan $\beta = |\alpha/\omega|$, since they were not interested in such. Mr. Koenig's mentioned curves for the cubic are plotted with the aforementioned K and r parameters as variables, hence are not directly applicable when a_1, a_2, and a_3 are the variables of interest. If desired, a set of tan β curves for direct use with these coefficients can be formulated from our curves. Or a very useful set of curves for $a_1 = 1$ can be obtained by replotting Mr. Koenig's curves through $a_2 = K$ and $a_3 = K/(1+r)$, as he remarks.

Mr. Koenig is correct in his statement that the sets of A and A' curves yield all the real roots. For if α in equation 8 of our paper takes on the values γ, δ, and ϵ in turn, it then follows from Figure 2(A) of the paper that equation 9 will have a zero root, whence perforce $b_3 = 0$, which is the condition yielding equation 13 of the curves in question. We regret this oversight on our part and very much appreciate having Mr. Koenig direct our attention to it.

Mr. Koenig's comment on our statement following equation 19 is very well made. In formulating this we had in mind Sponder's result that if a real-coefficient equation has one or more pairs of pure imaginary roots, then necessarily the Hurwitzian determinant Δ_{n-1} must vanish. This of course, as worked out by Sponder, is a necessary condition and, as Mr. Koenig indicates, not a sufficient condition. But when $a_3 \neq 0$, the authors are of the impression that the sufficiency thereof follows from the analysis associated with theorem C of Wall's[9] paper; hence they query Mr. Koenig's remark that "they are mistaken."

The authors are indeed indebted to Mr. Koenig for his penetrating and perspicacious discussion, which admirably elucidates certain points of this paper.

References

1. Ueber Directwirkende Regulatoren, J. Wischnegradski. *Der Civil-Ingenieur*, volume 23, 1877, pages 95–131.

2. See reference 6 of the paper.

3. See reference 7 of the paper.

4. R. Grammel. *Ingenieur-Archiv* (Berlin, Germany), volume 12, 1941, page 175.

5. See reference 9 of the paper.

6. Stability of the Parachute and Helicopter, H. Bateman. *Report Number 80, Fifth Annual Report*, National Advisory Committee for Aeronautics, (Washington, D.C.), 1919, pages 788–89.

7. A Treatise on the Stability of a Given State of Motion (book), E. J. Routh. Macmillan and Company (London, England), 1877, chapter III.

8. Zur Darstellung des Stabilitatsgebietes bei Schwingungsaufgaben mit Hilfe der Hurwitz-Determinanten, E. Sponder, *Schweiser Anchiv fur angewnadte Wissenschaft und Technik* (Solothurn, Switzerland), volume 16, 1950, pages 93–96.

9. Polynomials Whose Zeros Have Negative Real Parts, H. S. Wall. *American Mathematical Monthly* (New York, N. Y.), volume 52, 1945, pages 308–22.